EXPLORATIONS IN ANALYSIS OF STOP CONSONANTS USING WAVELET TRANSFORM

A thesis submitted

in partial fulfilment of the requirements

for the degree of

Master of Technology

by

Md Shazi Hashmi

to the

Department of Electrical Engineering

Indian Institute of Technology , Kanpur

March , 1996

This is to certify that the work titled **Explorations in analysis of stop consonants using Wavelet Transform** has been carried out by **Md. Shazi Hashmi** , Roll number **9410429** under my supervision and it has not been submitted elsewhere for a degree

Dr. S . K . Mullick
Prof Dept of EE
I I T Kanpur, India
(Thesis supervisor)

9/4/96

EE-1996-M-HAS-EXP

Abstract

We present some explorations in the analysis of six stop consonants /k, p, t, b, d, g/ using wavelet transform domain information in their acoustic manifestation Three different strategies are tried, they are analysis via classification, waveletogram and reconstruction from modulus maxima The main difficulties of the stop consonant problem lie in the *nonstationary* and *nonlinear* statistical structure of the acoustic signal in the *burst* and transition regions Nonstationarity renders the application of the Fourier Transform(FT) methods questionable Wavelet Transform has demonstrated good-time frequency localization properties and is therefore appropriate tool for the analysis of non–stationary signals like speech Moreover, unlike LPC and HMM modeling, we do not assume here any model for input speech The Discrete Wavelet Transform(DWT) may also be implemented as fast, pyramidal algorithm The analysis via classification produces 83% correct classification for unvoiced stop consonants /k, p, t/ We were looking for some explicit time information like *voice onset time* , place of occurrence of burst etc from the waveletogram but it fails to give explicit result Mallat's algorithm for signal reconstruction from the value of modulus maxima and their positions is successfully implemented in an effort to characterize the signal in terms of these features

Contents

1 Introduction **2**

1 1 Review 2

1 2 Organisation of the Thesis 5

2 Wavelet Transform **6**

2 1 Introduction 6

2 2 Discrete Wavelet Transform 8

2 3 Multiresolution Analysis(MRA) 10

2 4 Scaling Function and Wavelet Obtained from Iterated Filters 13

2 5 Fast Wavelet Transform 16

3 Methods of Analysis **18**

3 1 Analysis Through Classification . 18

3 1 1 Introduction . . 18

3 1 2 Classification Method 19

3 2 Waveletogram 20

3 3 Reconstruction from Multiscale edges . . 21

3 3 1 Introduction 21

3 3 2 Reconstruction Algorithm . 21

3 3 3 Numerical Reconstruction of 1-D Signals from local Maxima 24

4 Results and Discussions **25**

4 1 Data Set and the Experiment . . 25

4 2 Feature Extraction from Waveletogram 26

4 3 Simulation Results of Reconstruction Algorithm 27

5 Conclusion and Suggestion for Future Work 47

List of Figures

2 1 The Dyadic Sampling Grid 9

2 2 Decomposition of α_{j-1} into α_j and d_j 12

2 3 Reconstruction of α_{j-1} from α_j and d_j 12

2 4 Scaling Function for Daub-4 filter 14

2 5 Wavelet for Daub-4 filter 15

2 6 Scaling Function for Daub-10 filter 15

2 7 Wavelet for Daub-10 filter 16

2 8 Pyramidal algorithm for Analysis 17

2 9 Pyramidal algorithm for Synthesis 17

3 1 Alternate projection on space V and space Γ 24

4 1 Waveletogram of K 28

4 2 Waveletogram of P 29

4 3 Waveletogram of T 30

4 4 Waveletogram of B 31

4 5 Waveletogram of D 32

4 6 Waveletogram of G 33

4 7 Modified Waveletogram of K 34

4 8 Modified Waveletogram of P 35

4 9 Modified Waveletogram of T 36

4 10 Modified Waveletogram of B 37

4 11 Modified Waveletogram of D 38

4 12 Modified Waveletogram of G 39

4 13 Original Signal 40

4 14 Signal Reconstructed after 20 Iterations 41
4 15 Signal Reconstructed after 50 Iterations 42
4 16 Comparison of Original Signal and Reconstructed signal after 50 Iterations 43
4 17 Time plot of /k/ 44
4 18 Signal Reconstructed after 20 iterations 45
4 19 Signal Reconstructed after 50 iterations 46

Chapter 1

Introduction

1.1 Review

Speech is perhaps the most natural and fastest mode of not only human communication but also man-machine communication Successful developement of automatic speech recognition and synthesis systems for man–machine communication interface , is language dependent Different languages are characterized by different types and number of basic speech units , such as phoneme , different rules of pronunciation and joining of phonemes etc English language has 42 phonemes Out of that six stop consonants /k,p t,b,d,g/ present the greatest diffuculty in recognition tasks An efficient compact characterization of these stop consonants from their acoustic manifestation is therefore crucial step in the developement effort for speech recognition and synthesis systems This is the major objective of the present exploration

The methodology for achieving the above mentioned objective is to analyze the acoustic signal corresponding to the utterances of stop consonants using appropriate signal processing techniques and in particular the so called "*Wavelet Transform*" The rationale for this approach is elaborated in the following discussion

Throughout the history of digital speech processing , which is almost four decades old at present , the effort has been on to find accurate representations of speech signals that can be used for speech coding , enhancement , recognition and synthesis Drawing analogy

from the uniformity (whithin limits) of processing units in the human brain , and being driven by implementational considerations , researchers have tried to seek out flexible structures that comply with the vast bodies of loosely–related empirical results Following rapid advances in the science of spectral estimation in the past two decades , it has been possible to find reasonable accurate linear models for the stationary parts of vowel spectra However these , models (including Fourier transform with windowing , STFT and LPC) fails in terms of their transient–detection capability and scalability The result is that it is not possible to extend their use to the characterization of stop consonants

Stop consonants are acoustically characterized by rapid changes in short–time energy spectrum preceded or followed by a fairly long period (of the order of several centiseconds) during which there is no energy in all bands about the voicing component (i e above 300 Hz) The rapid opening of the oral cavity during the articulation of stops produces a wave–burst with a high–frequency spectral content Further , when a stop is adjacent to a vowel , the movement of the oral cavity to/and or from the closure results in rapid changes in the formant frequencies , known as transitions Since all these events occur within a fraction of the time taken for vowel articulation , a large or flexible window size will lead to the loss of related acoustic cues

From the psysiological point of view , six stop consonants /k,p,t,b,d,g/ , differ from one another in the phonetic features , the place of *articulation* and the phonetic feature *voicing* The former divides the six stop consonants into *labial* /p,b/ (closure of the vocal tract at lips) , *aleveolars* /t,d/ (closure occurs behind the teeth) and *velars* /k,g/ (closure occurs at the velam) The feature voicing divides the consonants into *voiced* /b,d,g/ and *voiceless* /p,k,t/ consonants Physiologically voicing is related to the timing between the larynx's muscle activity and the vocal tract articulation In non–whispered speech , voicing is represented by so called *voice onset time* (VOT) or *phonation onset parameter* defined to be the time interval between the release of the burst and onset of vocal cords pulsing (i.e onset of phonation)[9]

There is a greater statistically variability associated with VOT , depending upon , for example , the phonetic environment (e g a stop consonant may be located between two vowels or it may be word–initial or syllable final consonant) Furthermore , the voiced stop consonants /d,b,g/ may be either prevoiced (negative VOT or "voicing lead ") , in the sense that the vocal cords pulsing begins 5 to 15 ms after the burst release The

voiceless stop consonants /p,k,t/ have long VOT (vocal cords pulsing begins 20 to 40 ms after the burst release) The existence of invariant acoustic features for place of articulations is easier to understand than that for phonetic feature voicing Consideration of the vocal tract articulatory activity indicate that such features for place of articulation should be located in the sense of milliseconds in the vicinity of the burst release During this time interval , the articulations have not yet moved into the target positions for the next speech segment and consequently are less likely to be affected by the transition from one articulatory position to another

However the wavelet transform , a relatively recent developement in signal processing may provide a solution to some of these problems Frozen at an instant of time , the wavelet coefficients represent the instantaneous time vector output of a constant–Q filter bank ; for a particular filter , the output over time is a band–pass filtered , decimated time series

One of the important characteristics of the wavelet transform is the inverse variation of the analysis windows with the centre frequencies of the respective band–pass filters This may serve as an advantage in analyzing short–time transition phenomena like stop consonants The fact that the outputs for a particular scale form a time series may allow the methodologies that make use of time domain features like phase characteristic or time synchronization information On the other hand , an easy extension allows the interpretation of wavelets as as filter–bank coefficients At the same time , it might be noted , that appropriate windows will automatically be chosen for analysis of lower–frequency formants of accompanying vowels , thereby benefitting from the flexibility of the approach

The scalability of model derives its character from its interpretation as the projection onto a series of nested functional subspaces at varying resolutions , thus explaining the use of wavelets for multiresolution analysis and signal compression

The implementation of a discrete wavelet transform involves the inner product with the scaled and shifted version of a single “mother ” wavelet , the mother wavelet therefore crucially defines the wavelet basis For the analysis of the speech , thus one major question to be solved is to find a mother wavelet that will allow the the efficient representation of speech signals For a specific choice of wavelet basis , the next question to be solved is how the wavelet coefficients correlate to stop consonants in general , and the respective position and manners of articulation (which uniquely characterize stop consonants) in particular With these questions in mind three methods of analysis are explored in the

present thesis

1.2 Organization of the Thesis

Wavelet Transform theory is described in Chapter 2 Chapter 3 describes the methods of analysis In this chapter three methods of analysis have been explored Results and Discussions are described in Chapter 4 Finally the Chapter 5 deals with the conclusion and future scope of work Appendix A describes the projection operator on the affine space Γ

Chapter 2

Wavelet Transform

2.1 Introduction

For one dimensional function $f(x)$, the fourier transform $F(\xi)$ is another representation of the same function specified in frequency domain The fourier transform (and subsequently the inverse fourier transform) is orthogonal and all information in the original signal is preserved in the frequency domain We can write the inverse expression in the form

$$f(x) = \int_{-\infty}^{+\infty} F(\xi) e^{-j2\pi\xi} d\xi \tag{2 1}$$

Where $F(\xi)$ is fourier transform of $f(x)$ given by

$$F(\xi) = \int_{-\infty}^{+\infty} f(x) e^{j2\pi\xi x} dx \tag{2 2}$$

This indicates that the function $f(x)$ is expressed in terms of its projection on the basis function $e^{j2\pi\xi x}$ These functions are perfectly localized in frequency domain , but have infinite extent in time domain Hence any time localized information in $f(x)$ such as abrupt changes is spread in whole spectrum Hence fourier transform lacks the knowledge about the time localized features of the signal To overcome this problem Gabor introduced the window fourier transform also known as "*Short Time Fourier Transform(STFT)*" The

STFT of a signal $f(t)$ is defined as

$$STFT_f(\tau,\xi) = \int_{-\infty}^{+\infty} f(t)h(t-\tau)e^{j2\pi\xi t}dt \tag{2 3}$$

where $h(t)$ is a low pass window like gaussian window, which is localized in frequency domain as well as in time domain The parameter ξ specifies the conventional frequency and τ specifies the time shift Thus the transform $STFT_f(\tau,\xi)$ has significant magnitude only if the signal $f(t)$ has significant spectral component of frequency ξ in the neighbourhood of the time instant $t=\tau$ It is shown that to get fine frequency resolution the window should be as wide as possible , and the condition of fine time resolution is that the window width should be as small as possible These conflicting requirements cannot be satisfied simultaneously and uncertainty principle gives the lower bounds on the product of time scale resolution and frequency scale resolution [1] The STFT uses a window of fixed support , hence of fixed bandwidth This type of analysis is called uniform filter bank analysis [3]

In this regard the Wavelet transform(WT) offers an advantage over the other time frequency distributions It allows one to have fine *time scale resolution* and vice versa Along with wavelets comes the notion of *scale* The *continuous Wavelet transform* of $f(x)$ is defined as

$$CWT_f(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(x)\psi(\frac{x-\tau}{a})dx \tag{2 4}$$

where the window function $\psi(x)$ is the *basic* or *mother wavelet* and is a bandpass function $\psi(\frac{x-\tau}{a})/\sqrt{a}$ is the *wavelet basis function* sometimes called as *baby wavelet* By a change of varaible $a\hat{x} = x$, the above equation becomes

$$CWT_f(a,\tau) = \sqrt{a}\int_{-\infty}^{+\infty} f(a\hat{x})\psi(\hat{x}-\frac{\tau}{a})d\hat{x} \tag{2 5}$$

showing the equivalence between scaling $\psi(x)$ in (2 4) or scaling $f(x)$ in (2 5) to obtain the Wavelet transform In Wavelet transform , the analyzing window is scaled (i e either dilated or contracted) and the shifting parameter τ is made a function of the scale parameter a . Thus the narrow windows are shifted by small interval and wider windows shifted by large intervals This results in " *constant Q* " filter bank analysis [3] The Wavelet transform given by eqn (2 4) and (2 5) is invertible provided the mother wavelet satisfies the *admissibility* condition [1] given as

$$\int_{-\infty}^{+\infty} \frac{|\Psi(\xi)|^2}{\xi} d\xi < \infty \tag{2 6}$$

where

$$\Psi(\xi) = \int_{-\infty}^{+\infty} \psi(x) e^{-j2\pi\xi x} dx \tag{2 7}$$

is the Fourier transform of $\psi(x)$
When $\psi(x)$ is having fast decay at infinity [1] , the admissibility condition can also be written as

$$\int_{-\infty}^{+\infty} \psi(x) dx = 0 \tag{2 8}$$

This indicates that $\psi(t)$ is oscillatory

2.2 Discrete Wavelet Transform

In both the eqns (2 4) and (2 5) , (a, τ) are continuous and there is redundancy in CWT representation of $f(x)$ For practical computation (a, τ) should take only a finite number of values (a, τ) are thus discretised on finite grid and given by

$$a = a_0^m, \ \tau = n\tau_0 a_0^m \tag{2 9}$$

where m,n are integers
The discrete wavelet transform of $f(t)$ is given by

$$DWT_f(m, n) = \int_{-\infty}^{+\infty} f(t)\psi_{mn}(t) dt \tag{2 10}$$

where

$$\psi_{mn}(t) = a_0^{-\frac{m}{2}} \psi(a_0^{-m} t - n\tau_0) \tag{2.11}$$

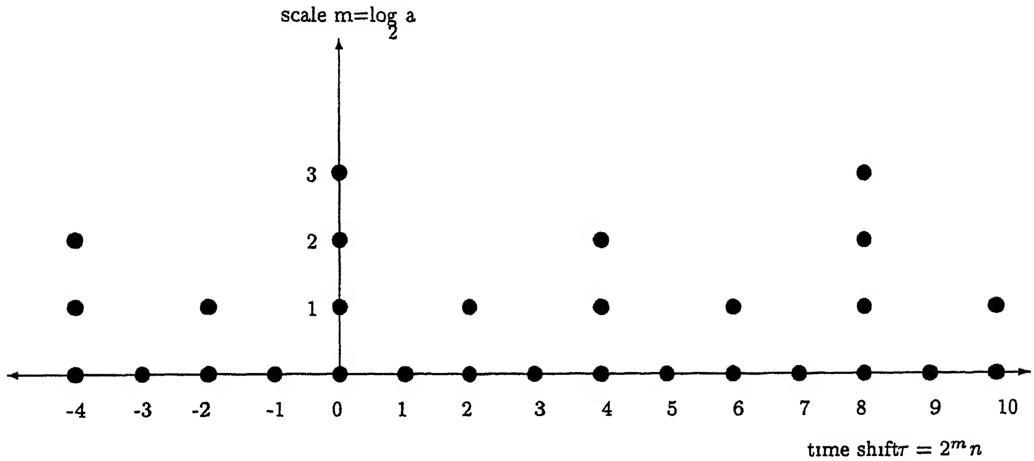

Fıgure 2 1: The Dyadıc Samplıng Grıd

$$\psi_{0,0}(t) = \psi(t)$$

and a_0 and τ_0 are constants that determıne the samplıng ıntervals

A practıcal samplıng scheme ıs $a = 2^m, t = n2^m$ ı e $a_0 = 2$ and $\tau_0 = 1$ so that eqn (2 11) becomes

$$\psi_{mn}(t) = 2^{-\frac{m}{2}}\psi(2^{-m}t - n) \tag{2 12}$$

Wıth thıs octave tıme and dyadıc translatıon , the sampled values of (a, τ) are shown ın dyadıc grıd of Fıg 2 1 Sınce the fourıer transform of $\psi(at)/\sqrt{a}$ ıs $\Psi(\omega/a)/a\sqrt{a}$, the centre frequency and bandwıdth of a wavelet are both scaled by $1/a$ for a tıme scalıng of a Thus the Q of all baby Wavelets

$$Q = \frac{Centrefrequency}{bandwıdth} = constant \tag{2 13}$$

gıvıng rıse to so called *constant Q analysıs* capabılıty of Wavelets The frequency resolutıon decreases with ıncrease ın centre frequencıes

2.3 Multiresolution Analysis(MRA)

The multiresolution analysis are widely used in computer vision for the purpose of pattern recognition In this section we will consider in brief , the relation between multiresolution analysis and wavelet transform A more detailed description can be found in [2]
Let $L^2(R)$ be the *Hilbert Space* of measurable , square -integrable 1-D functions f(x) The multiresolution analysis consists of breaking up $L^2(R)$ into ladder of spaces such that

$$\subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \tag{2 14}$$

with the properties that if $f(x) \in V_j$ then $f(x - 2^{-j}) \in V_j, k \in Z$ and $f(2x) \in V_{j-1}$ Let W_j be the orthogonal component of V_j in V_{j-1} This is written as

$$V_{j-1} = V_j \bigoplus W_j \tag{2 15}$$

The multiresolution analysis has the following other properties

1

$$f(x) \in V_j \Rightarrow f(2^j x) \in V_0 \tag{2 16}$$

2 There exists a *scaling function* $\phi(x)$ such that $\{\phi_{0,n}(x)\}_{n\in Z}$ is orthogonal basis of V_0 where $\phi_{0,n}(x) = \phi(x - n)$ This also implies that $\{\phi_{j,n} = 2^{-\frac{j}{2}}\phi(2^{-j}x - n)_{n\in Z}\}$ form the basis of V_j

3 As explained above , if W_j is the orthogonal component of V_j in V_{j-1} , then there exists a function $\psi(x)$ called as *wavelet* such that $\{\psi_{j,n}(x)\}$ is orthogonal basis of W_j where

$$\psi_{j,n}(x) = 2^{-\frac{j}{2}}\psi(2^{-j}x - n) \tag{2 17}$$

4 Let A_j denote the orthogonal projection operator on V_j i e the projection of $f(x)$ on V_j by $A_j f(x)$(also called as *approximate* signal at level j) Then $A_j f(x)$ is completely characterized by the sequence $\{\alpha_{j,n}\}$ where

$$\alpha_{j,n} =< f(x), \phi_{j,n}(x) > \tag{2 18}$$

5 The approximation at a resolution level j is a coarser approximation as compared to the approximation at resolution level $j-1$ The resulting loss of information is given by orthogonal projection of $f(x)$ on W_j and is called the *detail* signal at resolution level j Let D_j denote the orthogonal projection operator on the subspace W_j Then $D_j f(x)$ is completely characterized by the sequence $\{d_{j,n}\}$ where

$$d_{j,n} =< f(x), \psi_{j,n}(x) > \tag{2 19}$$

6 The function $\phi_{j,n}(x)$ and $\psi_{j,n}(x)$ satisfies the following relations

$$< \phi_{j,k}(x), \phi_{j,m}(x) >= \quad \delta_{k-m} \tag{2 20}$$

$$< \psi_{j,k}(x), \phi_{i,m}(x) >= \delta_{j-i}\delta_{k-m} \tag{2 21}$$

$$< \psi_{j,k}(x), \psi_{j,n}(x) >= \quad \delta_{k-n} \tag{2 22}$$

From the above discussion it is clear that $\{\psi_{j,n}\}_{j,n\in Z}$ is orthogonal basis of $L^2(R)$ Mallat has shown[2] how to obtain the sequence $\{\alpha_{j,n}\}$ and $\{d_{j,n}\}$ given the sequence $\{\alpha_{j-1,n}\}$ and vice versa Fig 2 2 shows the scheme for obtaining $\{\alpha_j\}$ and $\{d_j\}$ from $\{\alpha_{j-1}\}$ and Fig 2 3 shows the reverse scheme for obtaining $\{\alpha_{j-1}\}$ from $\{\alpha_j\}$ and $\{d_j\}$

The above two figures 2 2 and 2 3 are same as analysis and synthesis filter bank of 2 channel *quadrature mirror filter* (QMF) For perfect reconstruction G and H filters should satisfy the following conditions

$$|H(0)| = 1 \tag{2.23}$$

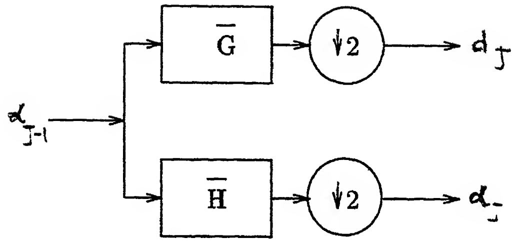

Figure 2 2 Decomposition of α_{j-1} into α_j and d_j

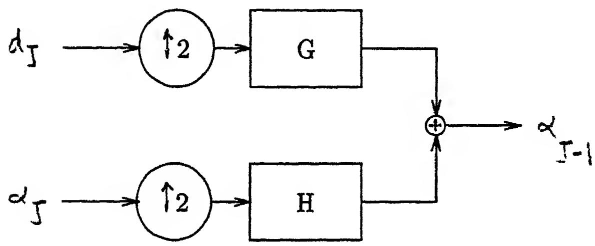

Figure 2 3 Reconstruction of α_{j-1} from α_j and d_j

$$|H(\omega)|^2 + |H(\omega+\pi)|^2 = 1 \tag{2 24}$$

$$H(Z)G(Z^{-1}) + H(-Z)G(-Z^{-1}) = 0 \tag{2 25}$$

where

$$H(Z) = \sum_{n \in z} h(n) z^{-n} \tag{2 26}$$

$$G(Z) = \sum_{n \in z} g(n) z^{-n} \tag{2 27}$$

$$H(\omega) = H(Z)|_{Z=e^{j\omega}} \tag{2 28}$$

$$G(\omega) = G(Z)|_{Z=e^{j\omega}} \quad (2\ 29)$$

$$\overline{H(Z)} = H(-Z^{-1}) \quad (2\ 30)$$

$$\overline{G(Z)} = G(-Z^{-1}) \quad (2\ 31)$$

The concept of multiresolution can be extended to 2-D signals [2]

2.4 Scaling Function and Wavelet Obtained from Iterated Filters

The scaling function $\phi(t)$ and wavelet $\psi(t)$ can be represented in terms of low pass and high pass filters [1] They are represented by

$$\phi(t) = \sqrt{2}\sum_{l} h(l)\phi(2t-l) \quad (2\ 32)$$

and

$$\psi(t) = \sqrt{2}\sum_{l} g(l)\phi(2t-l) \quad (2\ 33)$$

where g and h are low and high pass filters respectively , explained in section 2 3 Let p be the filter length Then the above eqns 2 32 and 2 33 can be written as

$$\phi(t) = \sqrt{2}\sum_{l=0}^{p-1} h(l)\phi(2t-l) \quad (2.34)$$

$$\psi(t) = \sqrt{2}\sum_{l=0}^{p-1} g(l)\phi(2t-l) \quad (2\ 35)$$

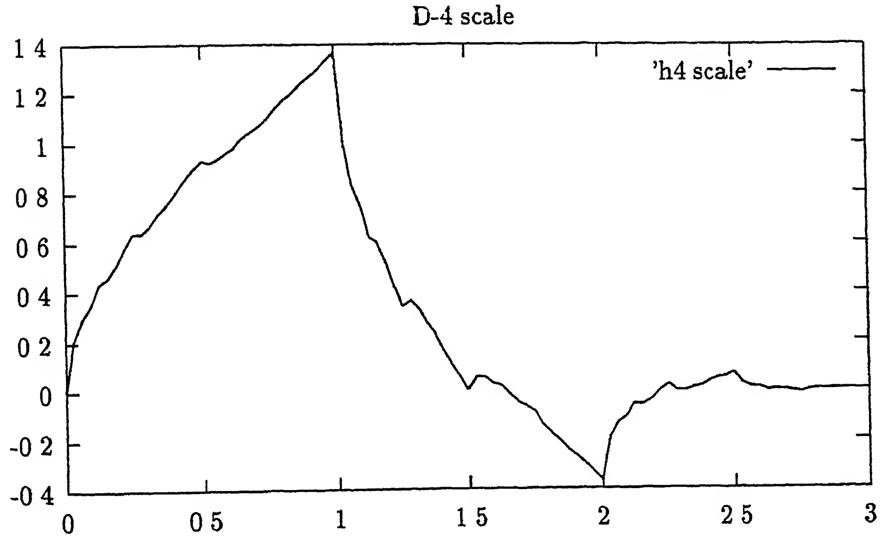

Figure 2 4 Scaling Function for Daub-4 filter

$\phi(t)$ is zero outside $t = 0$ and $t = p - 1$ [1] and since it is a continuous function , it must be zero at $t = 0$ and $t = p - 1$ Hence for an integer j , $\phi(t)$ is nonzero only for $j \geq 1$ and $j \leq p - 2$ Eqns (2 34) and (2 35) constitute $p - 2$ linear equations Which can be solved iteratively to obtain $\phi(t)$ and $\psi(t)$

The scaling function and wavelet for different filter lengths of Daubechies filter [5] are shown in figures 2 4 to 2 7 The filter coefficients are given in [6] Figs 2 4 and 2 5 shows the scaling function and wavelet of filter length 4 Figs 2 6 and 2 7 shows the scaling function and wavelet of filter length 10

It is seen that wavelet and scaling function , in general are nonsymmetric This results in nonlinear phase filters in analysis and synthesis filter banks By adopting bi-orthogonal filters , the condition of linear phase can be satisfied However this is at the cost of using separate scaling function and wavelet in analysis and synthesis filter bank For more details refer[14]

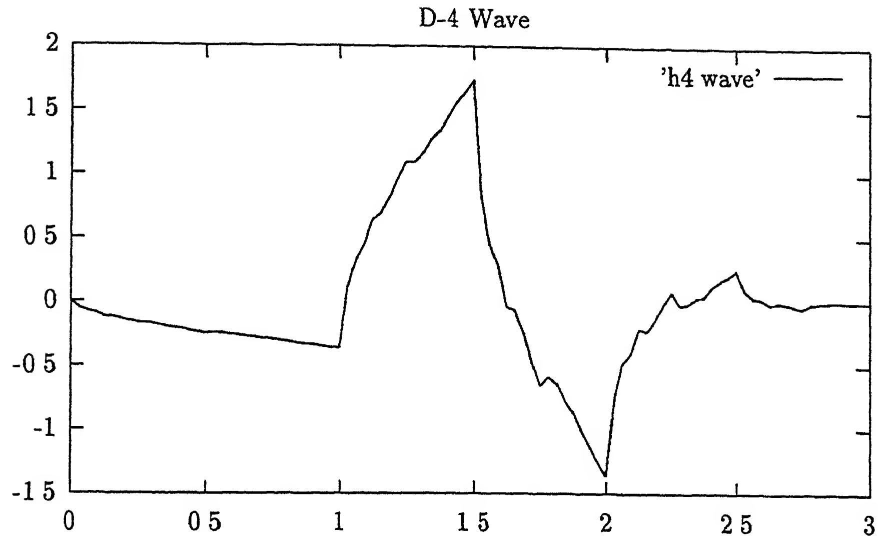

Figure 2 5 Wavelet for Daub-4 filter

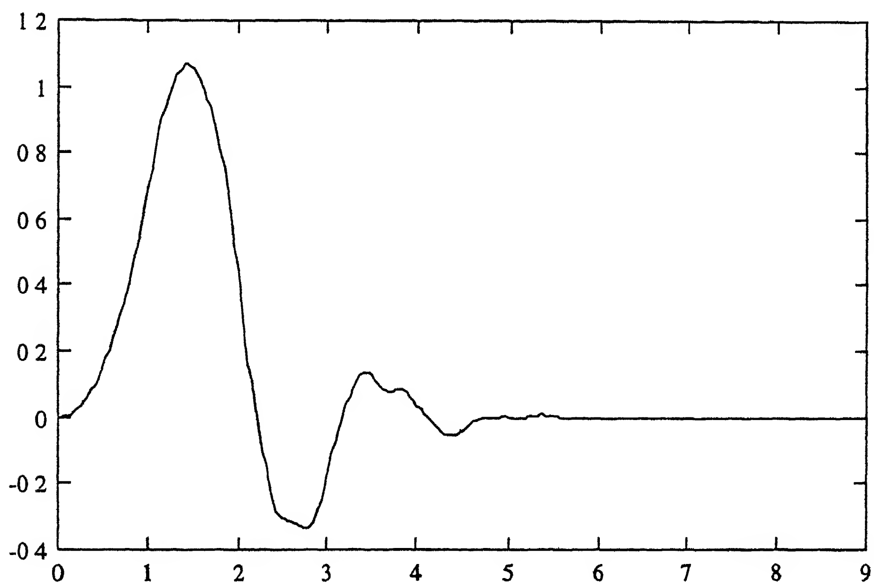

Figure 2 6 Scaling Function for Daub-10 filter

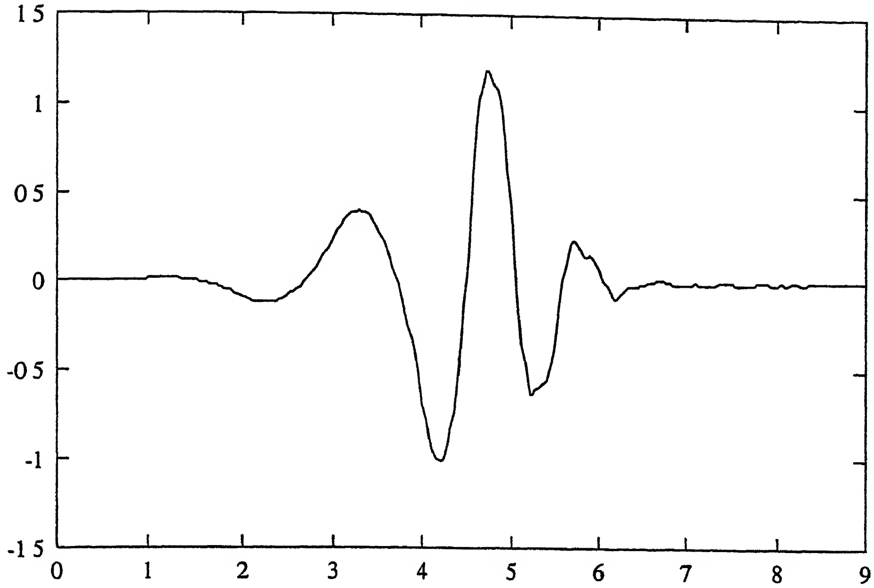

Figure 2 7 Wavelet for Daub-10 filter

2.5 Fast Wavelet Transform

In paractice the signal $f(x)$ is sampled with finite sampling rate $\frac{1}{T}$, to obtain the sequence $\{u(n)\}, u(n) = f(nT)$ These samples are supposed to be orthogonal projection of $f(x)$ on V_0 i e $\alpha_{0,n} = u(n)$ The wavelet coefficients $d_{j,k}, 1 \leq j \leq L$ can be obtained using the pyramidal algorithm as shown in Fig 2 8 Here L represents number of levels ($L \leq N$, where 2^N is the signal length)

The signal can be perfectly reconstructed from its wavelet coefficients The reconstruction scheme is shown in Fig 2 9

both Figs 2 8 and 2 9 results in fast wavelet transform

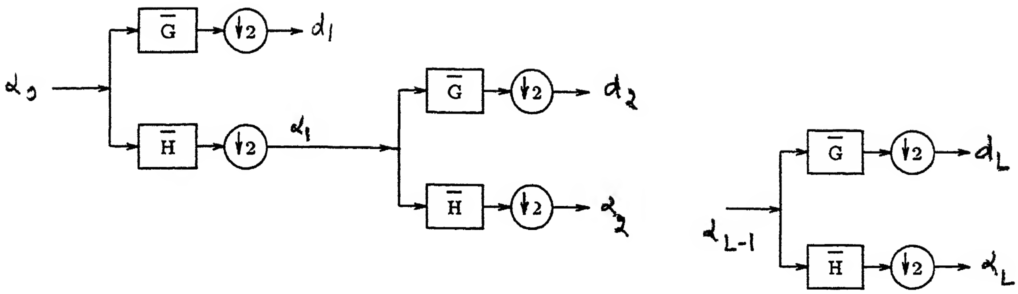

Figure 2 8 Pyramidal algorithm for Analysis

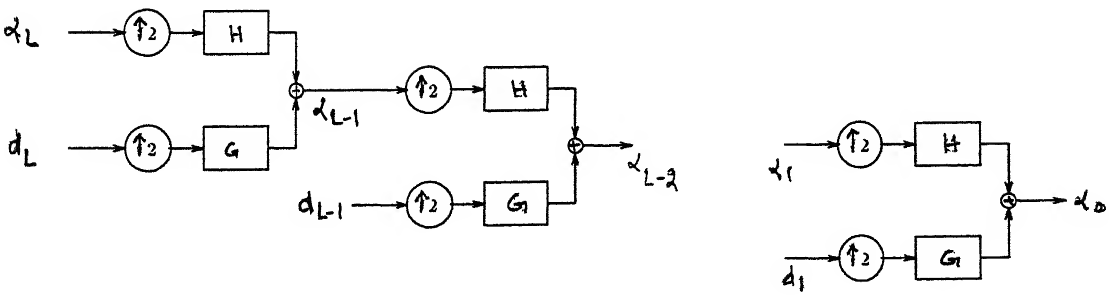

Figure 2 9 Pyramidal algorithm for Synthesis

Chapter 3

Methods of Analysis

Three diffrent methods have been explored for analysis purposes They are described in brief in the following sections

3.1 Analysis Through Classification

3.1.1 Introduction

Classification is attempted through a sequence of stored parameters that have previously been obtained through learning process Typically this process may be divided into two stages The first one is the feature extraction stage wherein short time temporal or spectral parameters of speech are extracted The second one is the classification stage wherein the derived parameters are compared with the stored reference parameters and decisions are made based on some kind of minimum distance rule We have used Discrete Wavelet Transform as the feature extraction tool Theory of wavelet transform has been explained in Chapter 2 With the wavelet transform we have used k-means[13] algorithm for finding the referance parameter

3.1.2 Classification Method

k-means algorithms comprise a very powerful class of clustering and segmentation methods[17 Variants of the basic algorithm have found use in a large variety of applications including compression and representation of speech data A prime example is the vector quantization(VQ) of speech signals

The training data are segmented into N clusters, each of which is represented by a reference vector These reference vectors are chosen such that the average of the distances between each element of the training data and its nearest reference vector is minimized Let $y_1, y_2, \quad , y_m \in R^m$ comprise of the training data set that we wish to segment into k clusters, where $2 \le K \le m$ we minimize

$$\varepsilon(W, R) = \sum_{i=1}^{k} \sum_{j=1}^{m} w_{ij} d(y_j, r_i) \tag{3 1}$$

subject to

$$\sum_{i=1}^{k} w_{ij} = 1 \quad j = 1, 2, 3, \quad , m$$

$w_{i,j} = 0 \;\; or \;\; 1 \quad i = 1, 2, \quad , k, j = 1, 2, \quad , m$

$R = [r_1, r_2, \quad , r_k] \in R^{nk}$

$W = [w_{ij}]$ is k $\times$ m matrix that defines the class membership $w_{ij} = 1$ indicates that $y_j \in cluster$ i, or more accurately, the cluster represented by r_i if

$$d(y_j, r_i) < d(y_j, r_l) \quad \forall l \neq i$$

d(,) is a distance measure and defined as $d(y_j, r_i) = (y_j - r_i)^T (y_j - r_i)$ R is the collection of the cluster centroids or reference vectors The problem in (3 1) is to determine the cluster memberships and cluster prototypes such that distortion error ε, is minimized Classification is performed by assigning the test token to the class of the nearest prototype For this reason, it is also called as a 1-nearest neighbour(1-NN) classifier

3.2 Waveletogram

As explained in the previous chapter , $d_{j,n}$ represents the wavelet coefficients and at level L , $d_{L,n}$ is the detail signal and $\alpha_{L,n}$ is the approximate signal For simplicity we set $d_{L+1,n} = \alpha_{L,n}$ and denote the wavelet coefficients by $\{d_{j,n}\}, 1 \leq j \leq L+1$,but $n = 0,1,2,\cdot \quad , 2^{N-j} for j \neq L+1$ and $n = 0,1,2, \quad , 2^{N-L} for j = L+1$ The two dimensional array d has a triangular structure

The plot of $\{|d_{j,n}|\}$ in the (j,n) plane corresponds to the spectrogram and is called Waveletogram [8] The usefullness of waveletogram lies in the fact that $|d_{j,n}|^2$ measures the energy of the signal at scale 2^j and location $2^{N-j}n$ For orthogonal basis $\hat{E} = \sum_{j,n} |d_{j,n}|^2$ is precisely the energy $E = \sum_{t=0}^{2^N-1} |f(t)|^2$ of the signal , where 2^N is the length of the signal Since the $2D$ array $d_{j,n}$ is triangular in nature , hence it is difficult to plot To overcome this difficulty we define a matrix of size $(L+1) \times (2^{N-1})$ or (level+1)$\times$(half the length of original signal) The elements of E-matrix are given by

$$E_{0,2^{L-1}k+n} = |d_{L+1,k}| \tag{3 2}$$

$k = 0,1,2, \quad , 2^{N-L}-1, n = 0,1,2, \quad \cdot, 2^L - 1$
for $j = L+1$

$$E_{L+1-j,2^{j-1}k+n} = |d_{j,k}| \tag{3 3}$$

$k = 0,1,2, \quad , 2^{N-j}-1, n = 0,1,2, \quad , 2^{j-1} - 1 for\ j = L, L-1, \quad 1$
The coefficients of E-matrix are mapped into 0 to 255 gray scale The resulting matrix plot is waveletogram which is used in our analysis

3.3 Reconstruction from Multiscale edges

3.3.1 Introduction

Points of sharp variations are often among the most important features for analyzing the properties of transient signals and images In Images , they are generally located at the boundaries of important image structures Signal sharp variation produces modulus maxima at different scales 2^j of its wavelet transform Mallat [4] has shown that the almost exact reproduction of the signal is possible from the modulus maxima of its wavelet transform That means a signal can be reconstructed back from the maxima information and place of their occurance at the dyadic scale 2^j Mallat's algorithm for reconstruction of 1-D signal is described in the next section It can also be extended to 2-D image [4]

3.3.2 Reconstruction Algorithm

Let $f(x) \in L^2(R)$ and $(W_{2^j} f(x))_{j \in Z}$ be its dyadic wavelet transform We describe an algorithm that reconstructs an approximation of $(W_{2^j} f(x))_{j \in Z}$, given the positions of the local maxima of $|W_{2^j} f(x)|$ and the values of $W_{2^j} f(x)$ at these locations For this purpose we characterize the set of functions $h(x)$ such that at each scale 2^j , the modulus maxima of $W_{2^j} h(x)$ are the same as the modulus maxima of $W_{2^j} f(x)$ Let $(x_n^j)_{n \in Z}$, be the abscissa where $|W_{2^j} f(x)|$ is locally maximum The maximum constraints on $W_{2^j} h(x)$ can be decomposed in two conditions

1 At each scale 2^j , for each local maxima located at x_n^j , $W_{2^j} h(x_n^j) = W_{2^j} f(x_n^j)$

2 At each scale , the local maxima of $|W_{2^j} h(x)|$ are located at the abscissa $(x_n^j)_{n \in Z}$

Condition (1) is equivalent to

$$< f(x), \psi_{2^j}(x_n^j - x) >=< h(u), \psi_{2^j}(x_n^j - u) > \qquad (3\ 4)$$

In general the function $h(x)$ that satisfies eqn(3 4) does not characterize f(x) uniquely[4] Condition (2) is more difficult to analyze because it is not convex In order to solve this problem numerically , we approximate condition 2 with a convex constraint Condition 2

defines the value of the wavelet transform at the point $(x_n^j)_{(j,n)\in Z^2}$ Instead of imposing that local maxima of $W_{2^j}h(x)$ are located at these points , we impose that $|W_{2^j}h(x)|^2$ be as small as possible on the average This generally creates local modulus maxima at the position $(x_n^j)_{(j,n)\in Z^2}$ The number of modulus maxima of $W_{2^j}f(x)$ depends on how much this function oscillates To have as few modulus maxima as possible outside the abscissa $(x_n^j)_{(j,n)\in Z^2}$, we also minimize the derivative of $W_{2^j}h(x)$ Since these conditions must be imposed at all scales 2^j , we minimize

$$||h||^2 = |W_{2^j}h(x)| = \sum_{j=-\infty}^{+\infty} (||W_{2^j}h||^2 + 2^{2j}||\frac{dW_{2^j}h}{dx}||^2) \tag{3 5}$$

Let us now describe an algorithm for our minimization problem Instead of computing the solution itself , we reconstruct its wavelet transform with an algorithm based on alternate projections Let K be the space of all sequences of functions $(g_j(x)_{j\in Z})$ such that

$$|(g_j(x))_{j\in Z}|^2 = \sum_{j=-\infty}^{+\infty} (||g_j||^2 + 2^{2j}||\frac{dg_j}{dx}||^2) < +\infty \tag{3 6}$$

The norm || || defines a Hilbert structure over K Let V be the space of all dyadic wavelet transforms of functions in $L^2(R)$ It can be proved that V is included in K Let Γ be the affine space of sequences of functions $(g_j(x)_{j\in Z}) \in K$ such that for any index j and all maxima positions x_n^j

$$g_j(x_n^j) = W_{2^j}f(x_n^j) \tag{3 7}$$

One can prove that Γ is closed in K The dyadic wavelet transform that satisfy condition (1) are sequences of functions that belong to

$$\Lambda = V \cap \Gamma \tag{3 8}$$

We must therefore find an element of Λ whose norm || || is minimum This is done by alternating projections on V and Λ

Dyadic wavelet transforms is invariant under operator[4]

$$P_V = W \circ W^{-1} \tag{3.9}$$

(where W^{-1} is the inverse wavelet transform)
For any sequence $X = (g_j(x))_{j \in Z} \in K$, it is clear that $P_V X \in V$, therefore, P_V is a projector on V. P_V is self-adjoint and orthogonal for a kind of wavelet (explained in chapter 2). The orthogonal projection on the space V is thus implemented by applying the operator W^{-1} followed by the operator W. P_Λ is the projection operator on affine set Λ which is orthogonal with respect to the norm $\|\ \|$. Mallat [4] has proved that the projection operator P_Λ can be implemented by adding piecewise exponential curves to each function of the sequence that we project on P_Λ. Appendix A characterizes the projection on the affine set Γ, which is orthogonal with respect to the norm $\|\ \|$. Let $P = P_V \circ P_\Lambda$ be the alternate projections on both spaces. Let $P^{(n)}$ be n iterations over the operator P. Since Λ is affine space and V a Hilbert space, a classical result on alternate projections [11] proves that for any sequence of functions $X = (g_j(x))_{j \in Z} \in K$

$$\lim_{n \to +\infty} P^n X = P_\Gamma X \tag{3.10}$$

Alternate projections on Λ and V converge strongly to the orthogonal projection on Γ. If X is the zero element of K which means that $g_j(X) = 0$ for all $j \in Z$, the alternate projections converge to the element of Γ, which is closest to zero, and thus whose norm $\|\ \|$ is minimum. Figure 3.1 illustrates how the approximation of the wavelet transform of $f(x)$ is reconstructed by alternating orthogonal projections on an affine set Γ and on space V of all dyadic wavelet transforms. The projections begin from the zero element and converge to its orthogonal projection on $\Gamma \cap V$

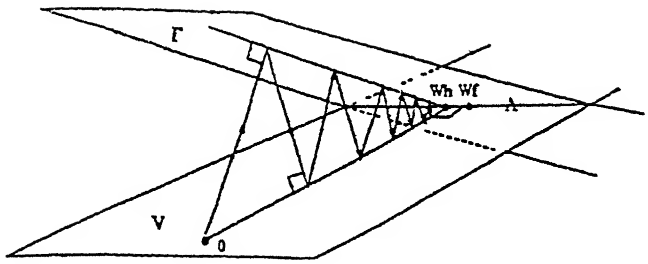

Figure 3 1 Alternate projection on space V and space Γ

3.3.3 Numerical Reconstruction of 1-D Signals from local Maxima

From the results of Meyers [12] , we know that in general , we cannot reconstruct exactly a function from the modulus maxima of its wavelet transform Mallat's algorithm approximates this inverse problem by replacing the maxima constraint by the minimization of a norm that yields a unique solution We thus do not converge toward the wavelet transform of the original signal but toward some other wavelet transform that we hope to be close to the original one The computation of solution might be unstable , in which case the alternate projections converge very slowly Hence more number of iterations are required for the solution close to the original signal

Chapter 4

Results and Discussions

4.1 Data Set and the Experiment

In order to study the features of stop consonants , data corresponding to three voiceless stop consonant /p, k, t/ were collected Data from a single male speaker as used as speech data base They consists of 19 /k/'s, 21 /t/'s, and 19 /p/'s, 59 in all These were next passed through a burst detector (a simple energy thresholding) and the first 100 ms after the burst were retained for further processing At 8 KHz sampling rate , they produce 800-long data records The wavelet transform of each data was then computed for number of levels $L = 5$ Five repetitions of each utterance were chosen at random for training set Ten templates per levels were obtained for each utterance using k-means algorithm In our case we have 50 templates for detail signal and 10 templates for the coarser signal at level 5 Classification is done through template matching (wavelet coefficients are compared with the stored templates level wise) minimum distance rule is used as method of classification The group of templates which produce minimum distortion is classifified to that group . Correct classification scores are shown in table 4 1 It is seen in the table that poorest classification is for /p/ and the overall classification is 83%(49/59) Higher scores are expected if *learning vector quantization*(LVQ) [16] as used instead of vector quantization Vector quantization minimizes the average distortion error and does not approximate the optimum decision boundaries between the different classes This drawback is taken care

/k/	/p/	/t/	Total
89 5%	73 7%	85 7%	83%
17/19	14/19	18/21	49/59

Table 4 1 Classification Score

of in LVQ

4.2 Feature Extraction from Waveletogram

Six stop consonants /k, p, t, b, d, g/ were spoken by a single male speaker in VCV syllables The waveletogram along with the time plot for each stop consonant is shown in Figs 4 1 to 4 6 We have tried to establish one to one correspondence between waveletogram and spectrogram, but did not succeed We can see in the figure that they are (waveletogram) certainly different from one another In fact we were looking for the explicit time information as we get in spectrogram We tried the modified waveletograaveraged visualization such as thresholding, smoothing, averaging etc The averaged waveletogram of each stop consonant is shown in Figs 4 7 to 4 12 In this trial also we could not get the explicit time information Observing the modified waveletogram of /k, g/, /p, b/ and /t, d/, we find some similarity between each pair This similarity is indicative of the same place of articulation of each pair

4.3 Simulation Results of Reconstruction Algorithm

The reconstruction algorithm of section 3 3 2 was tested on the data shown in Fig 4 13 The above data is the part of speech signal and is 64-point long Wavelet transform was applied on this data and extreme points (maxima and minima) and their places of occurance were calculated for each level The reconstruction algorithm was then applied on the resulting data and reconstructed signal are shown in Figs 4 14 and 4 15 Fig 4.14 is the plot of the signal obtained after 20 iterations The SNR is 25 dB/decade The signal obtained after 50 iterations is given in fig 4 15 looking at the Fig 4 15 and the original signal , one can hardly make any difference For easy comparison they are plotted on the same graph shown in Fig 4 16

The reconstruction algorithm was also applied to stop consonant /k/ with time plot given in Fig 4 17 Since the length of the speech data is very large (8000-long , recorded for 1 sec at sampling frequency 8 KHz) We apply the reconstruction algorithm frame wise We have chosen 256 as the frame length Figs 4 18 and 4 19 are the plots of reconstructed signal after 20 and 50 iteration's respectively

The reconstructed signals are not equal to the original signal but are numerically very close They have no spurious oscillations and have same types of sharp variations seen in Figs 4 17 and 4 19 Qualitatively , the reconstructed signals are thus very similar to the original one Error's are hardly noticeable by comparing the graphs shown in Fig 4 17 We have no upper bound on the error due to the distance between the signal to which we converge and the original signal This is an open mathematical problem , but the numerical precision of this reconstruction algorithm is sufficient for many signal processing applications like efficiency of representation (data compression) , perceptual relevance , and robustness to noise etc

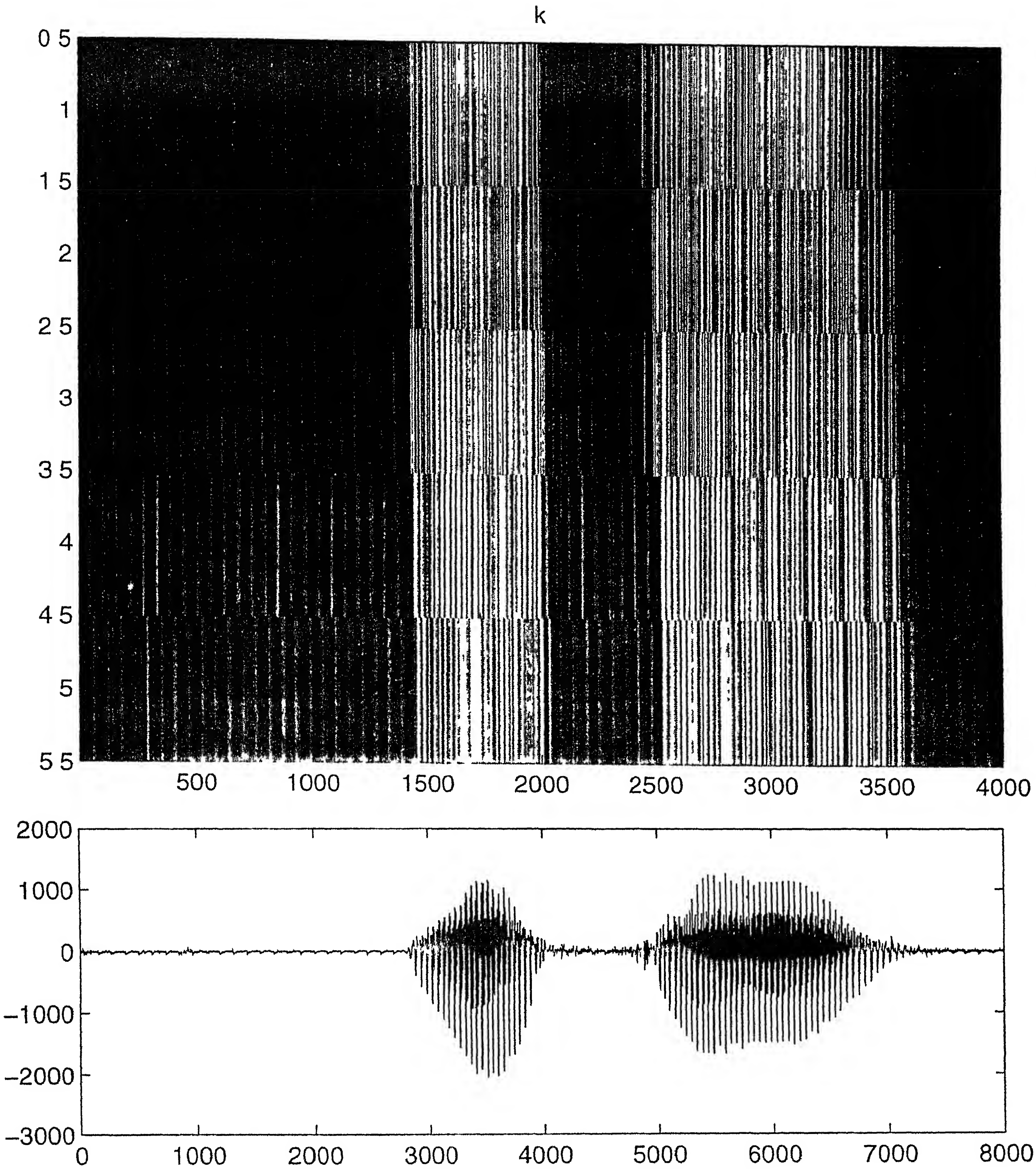

Figure 4 1 Waveletogram of K

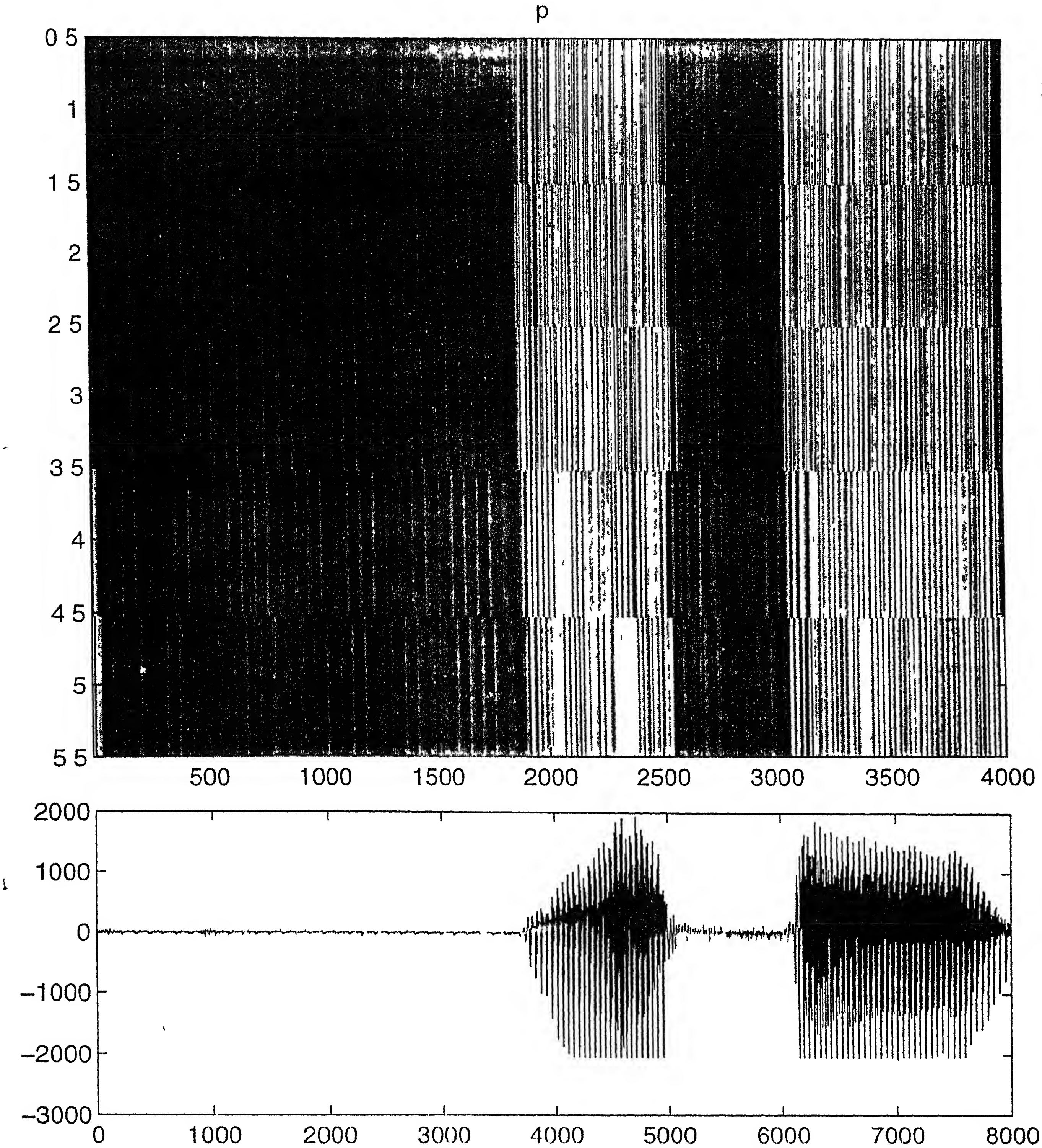

Figure 4 2 Waveletogram of P

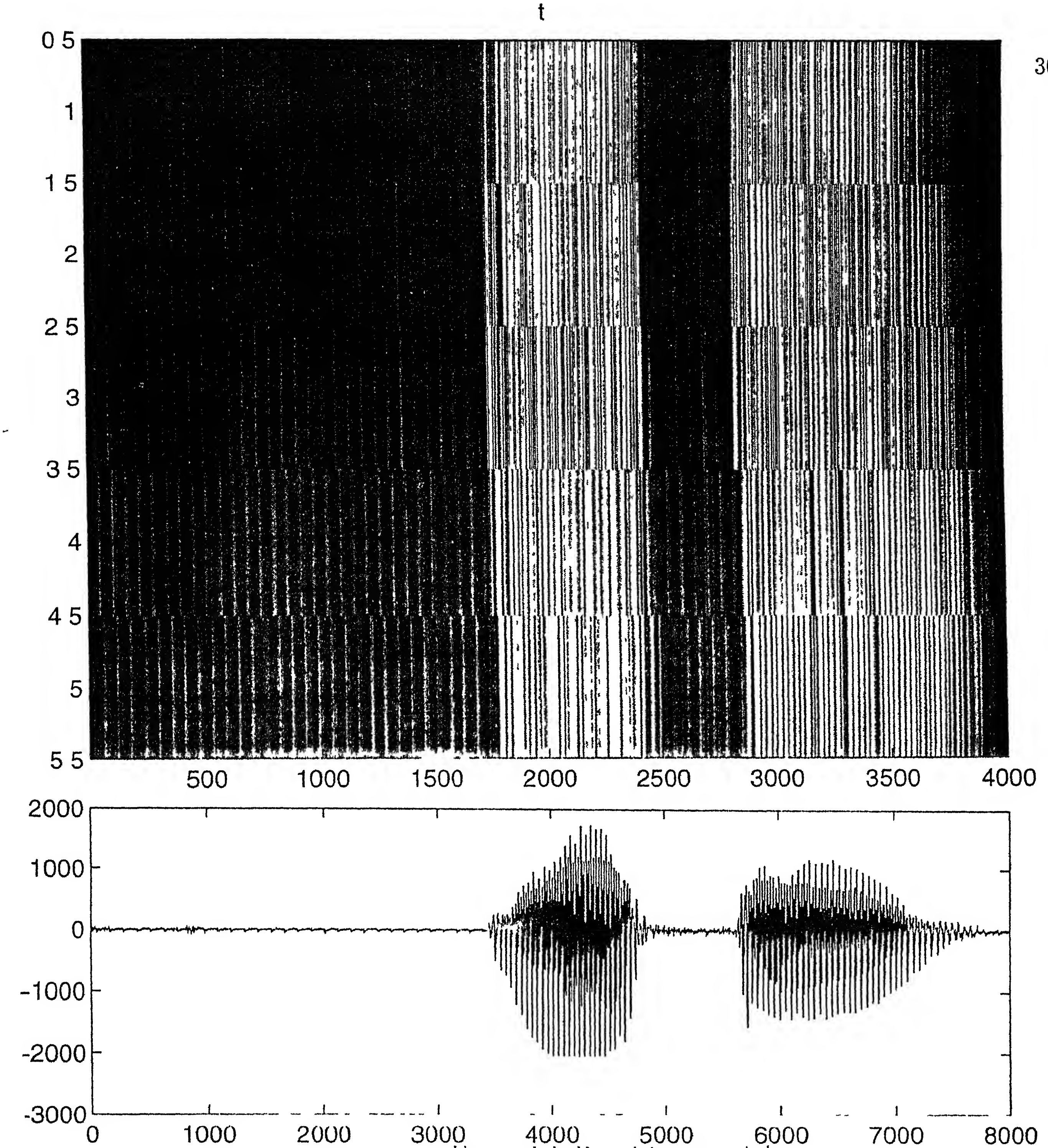

Figure 1 3 Waveletogram of 1

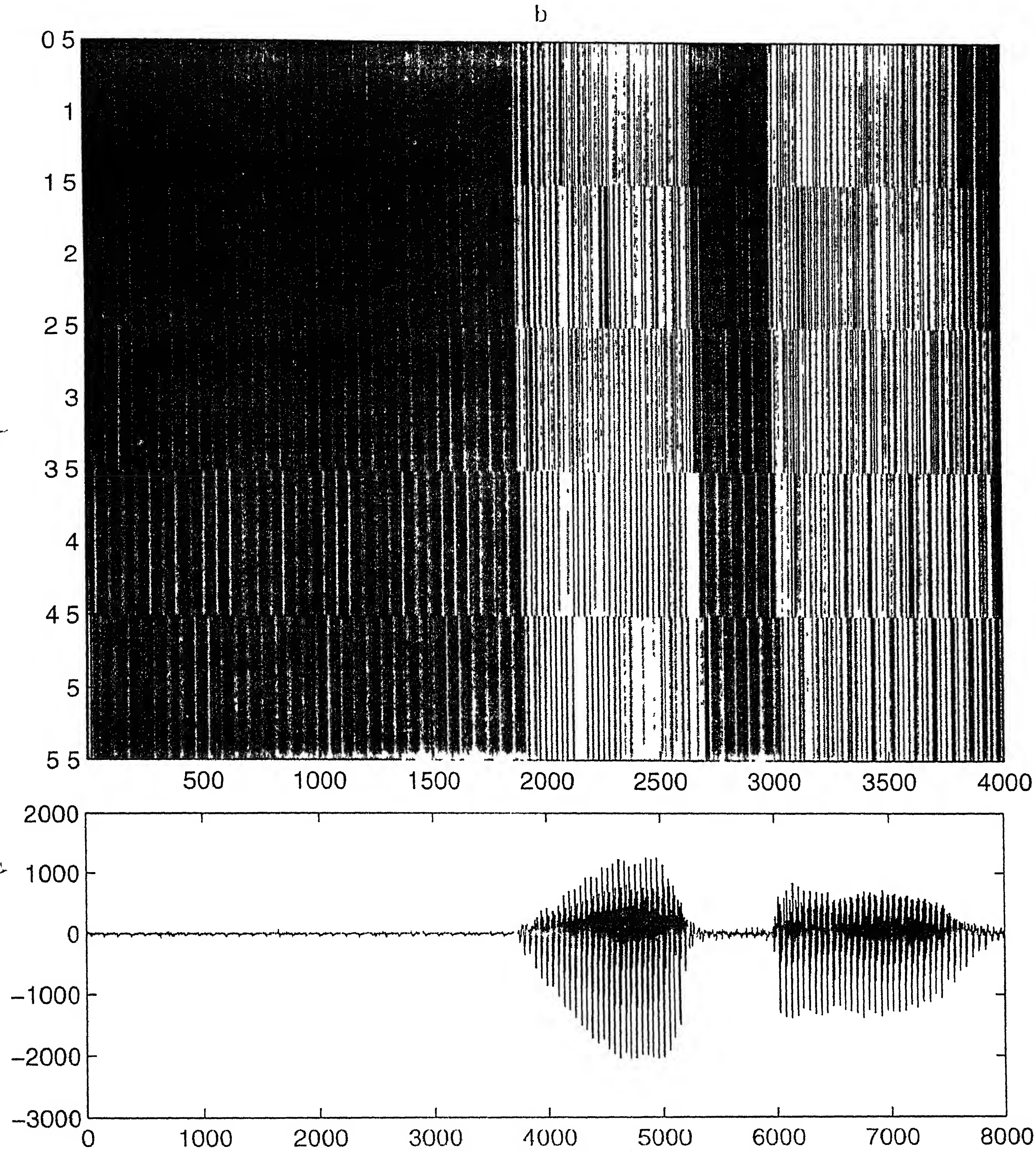

Figure 1 1 Waveletogram of B

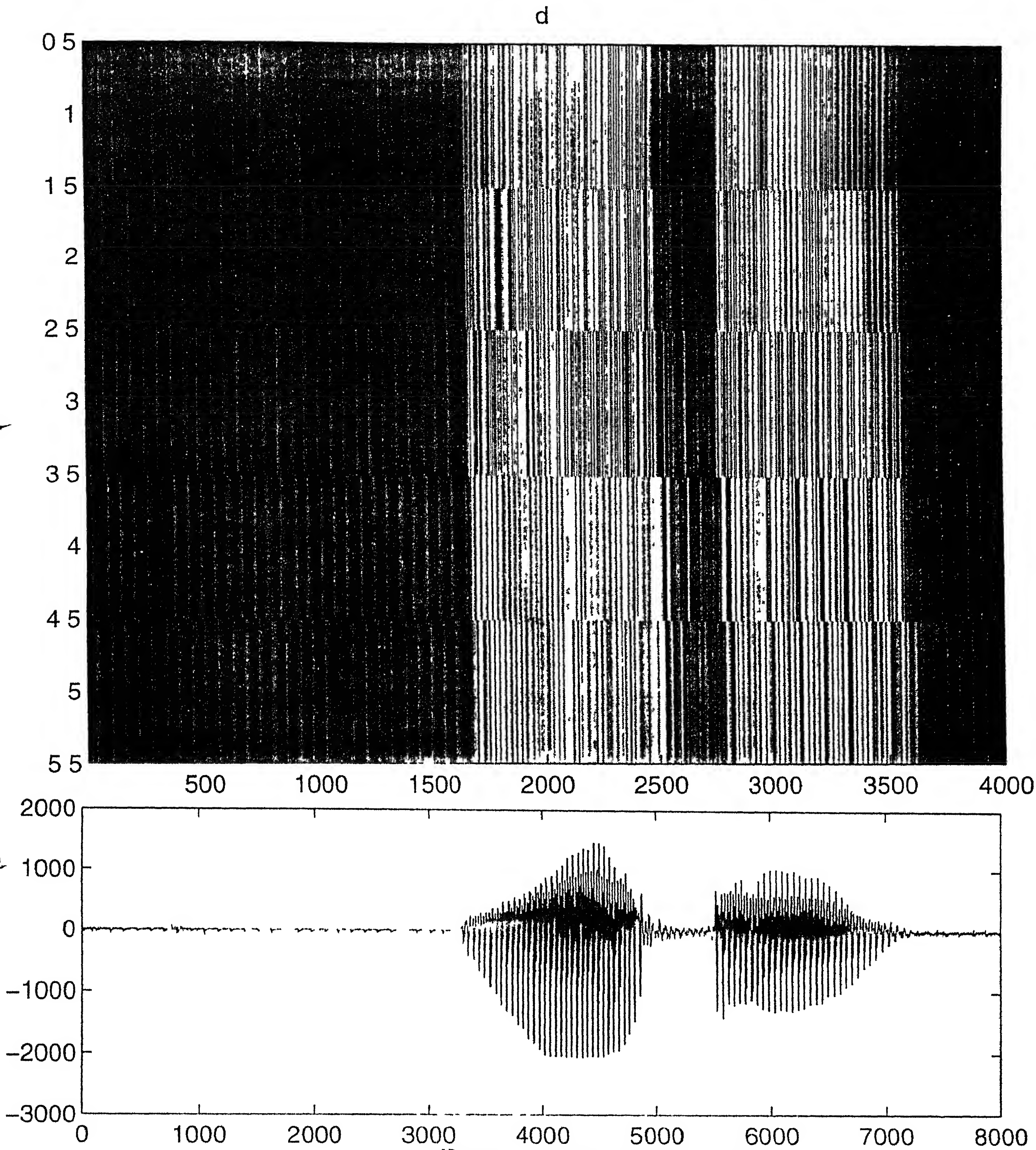

Figure 1 5 Waveletogram of D

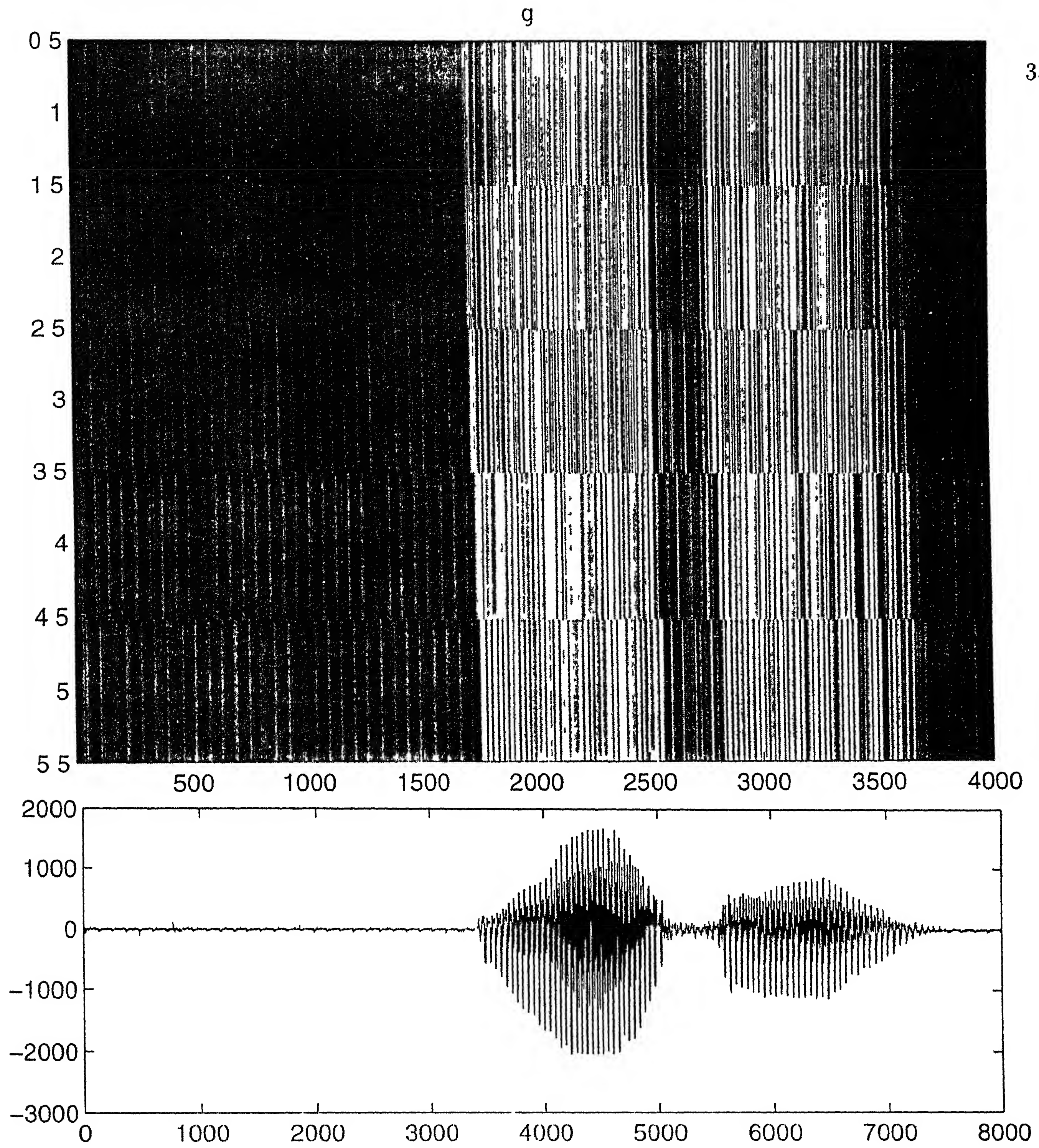

Figure 4 6 Waveletogram of G

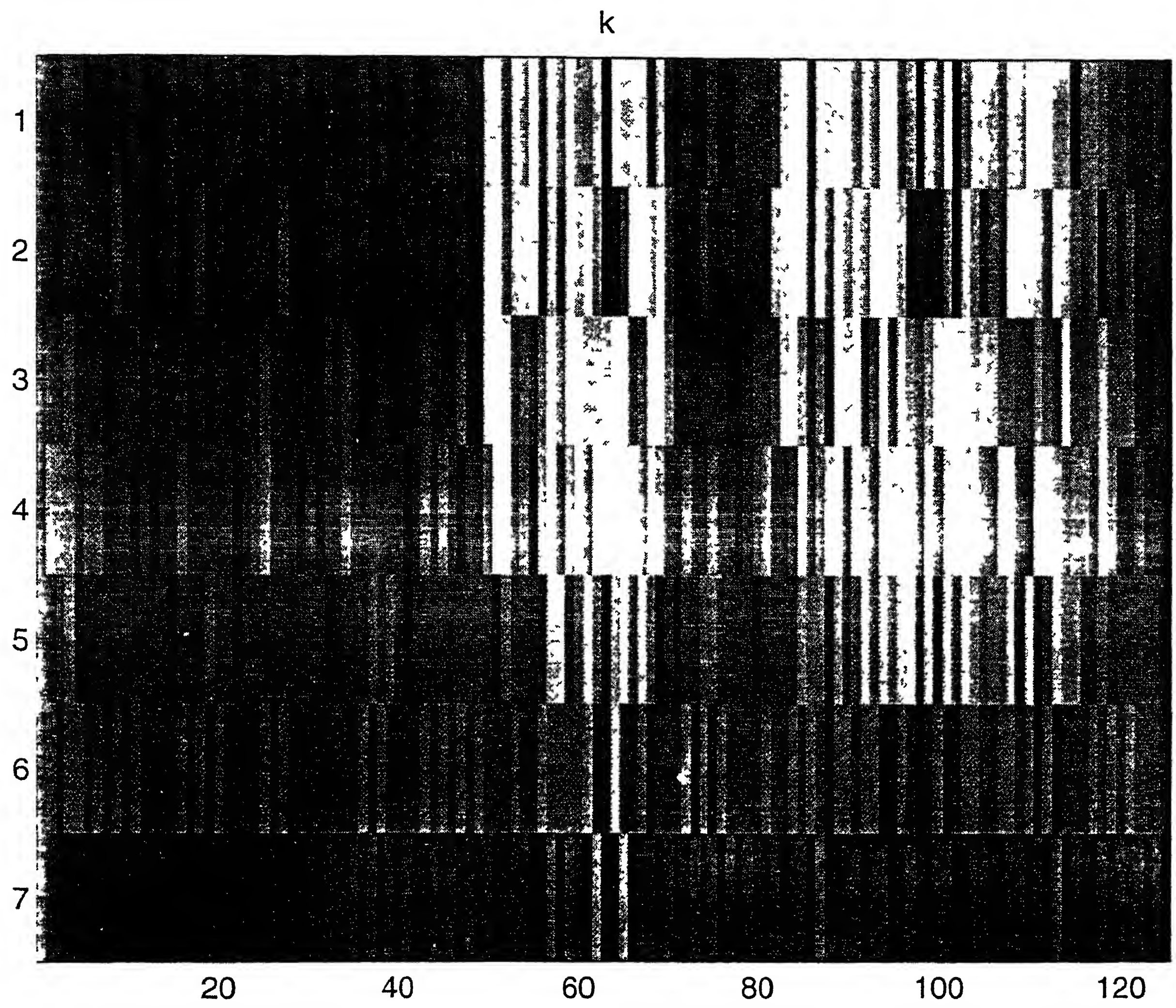

Figure 4 7 Modified Waveletogram of K

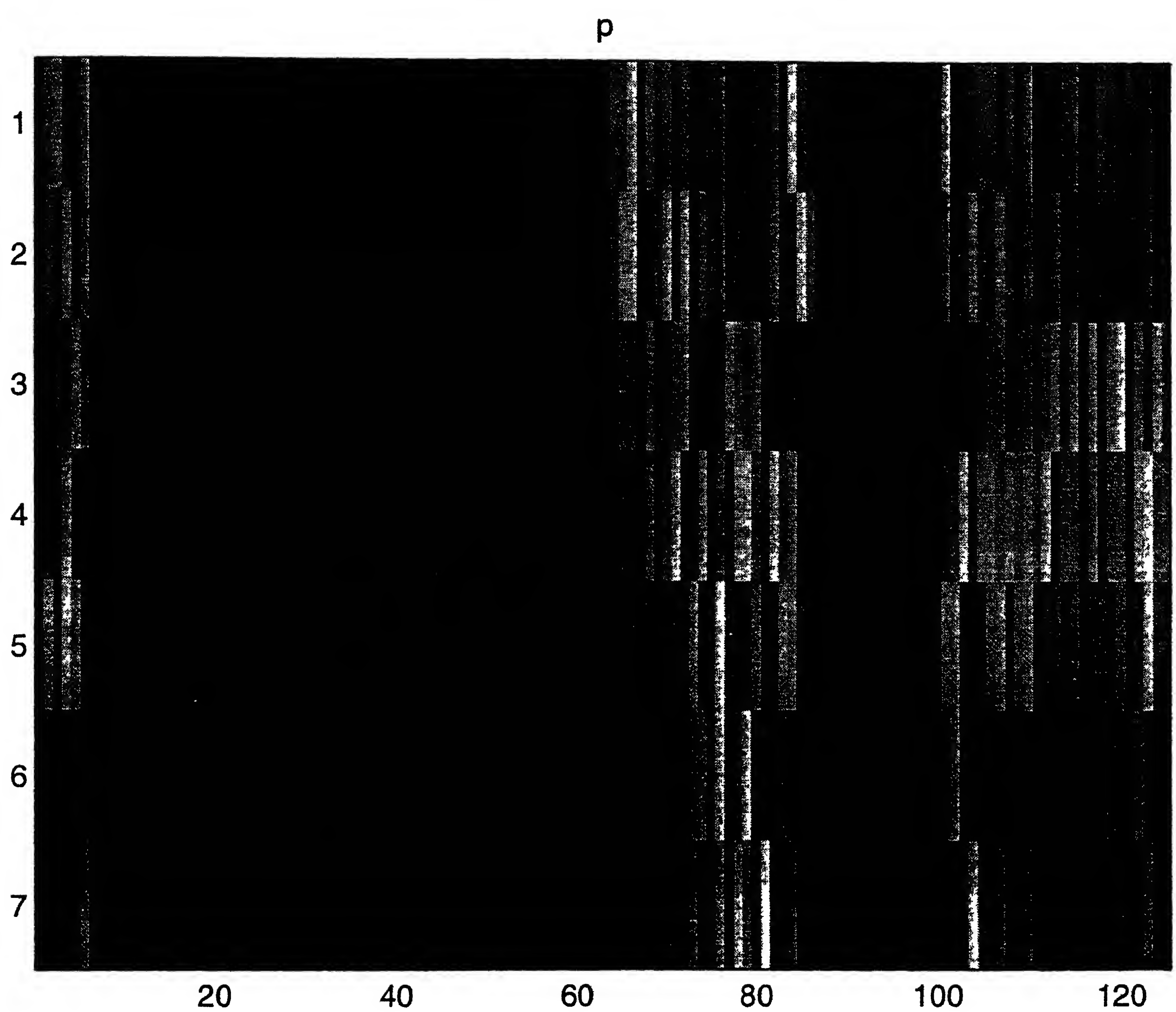

Figure 4 8 Modified Waveletogram of P

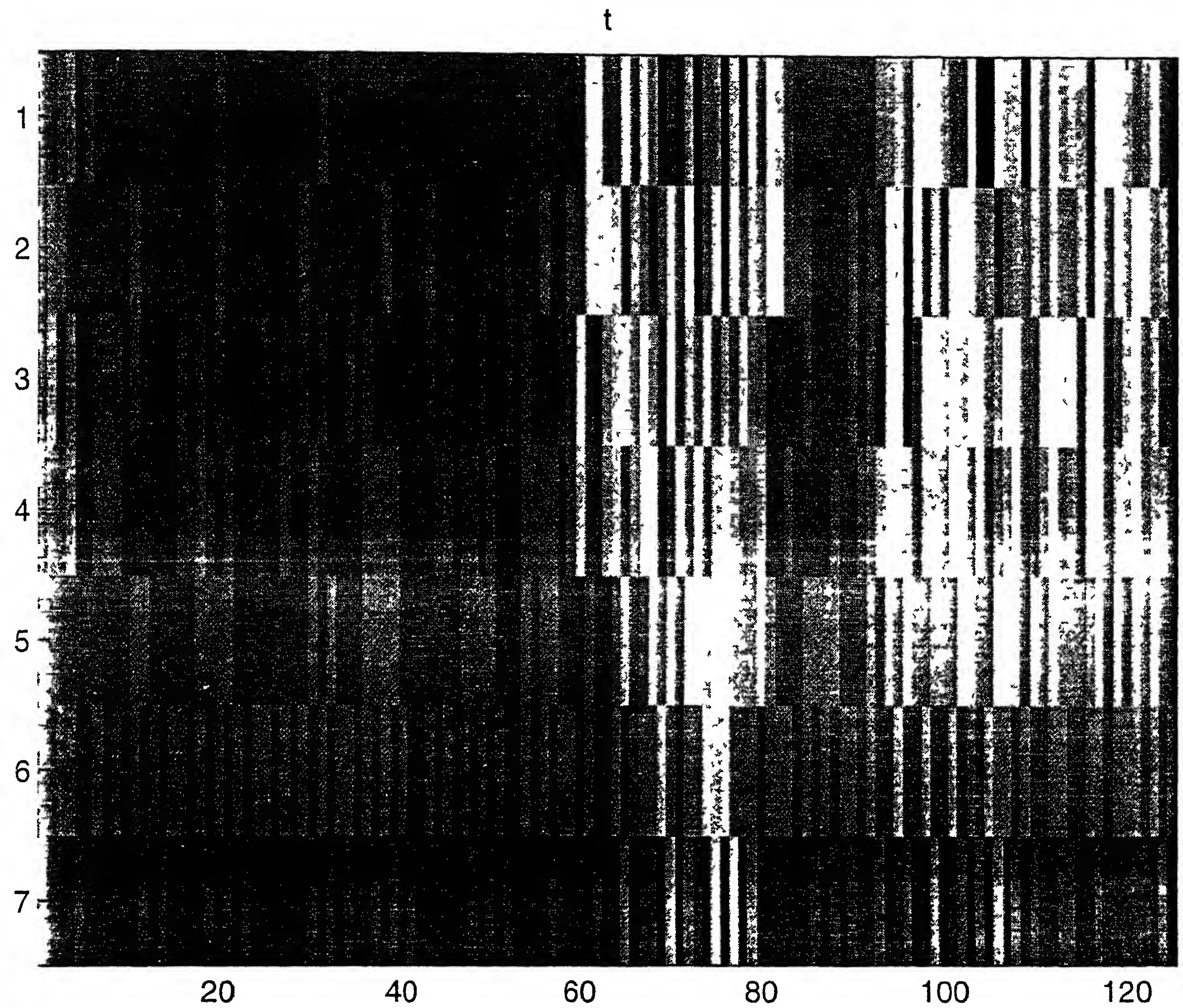

Figure 4 9 Modified Waveletogram of T

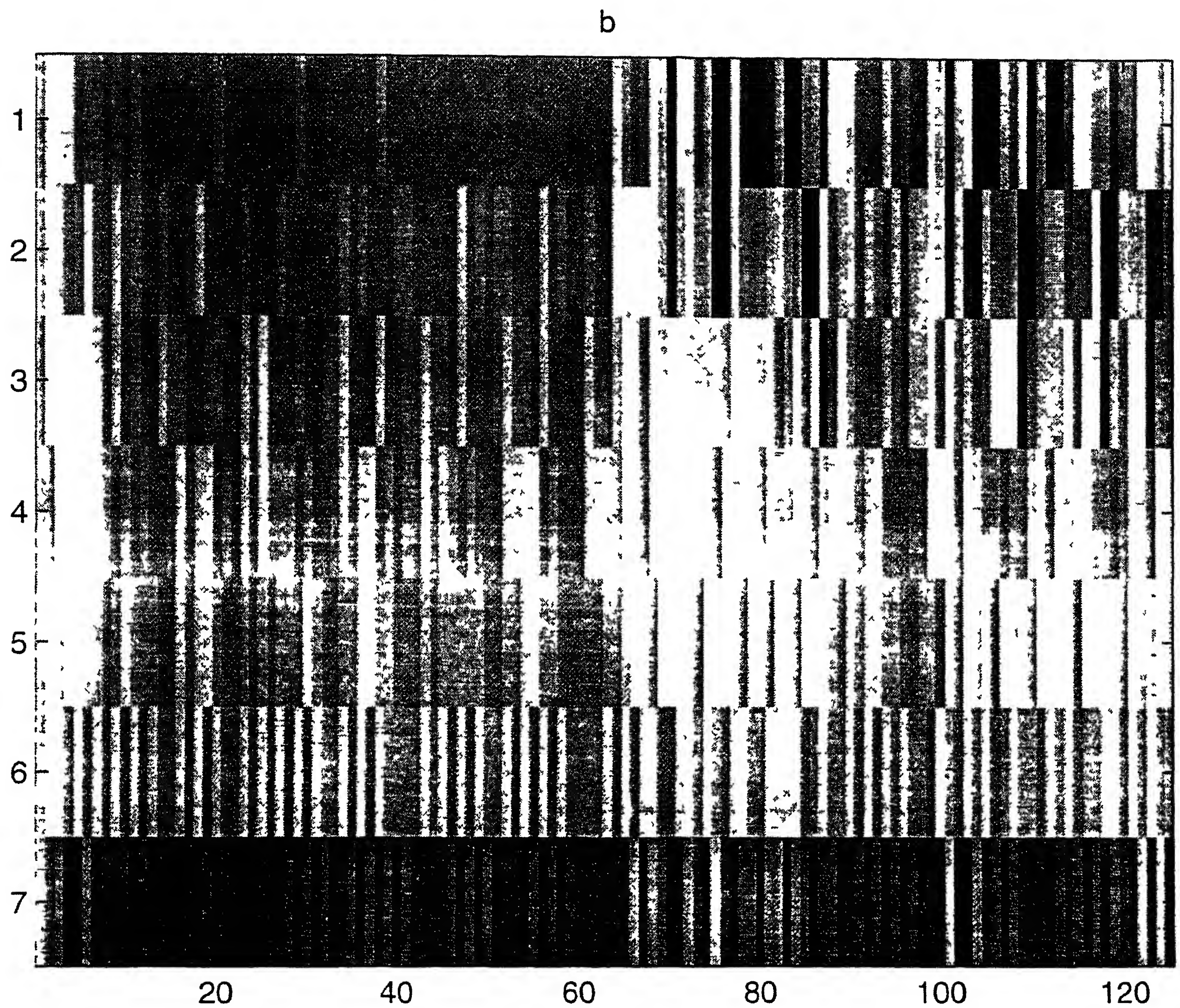

Figure 4 10 Modified Waveletogram of B

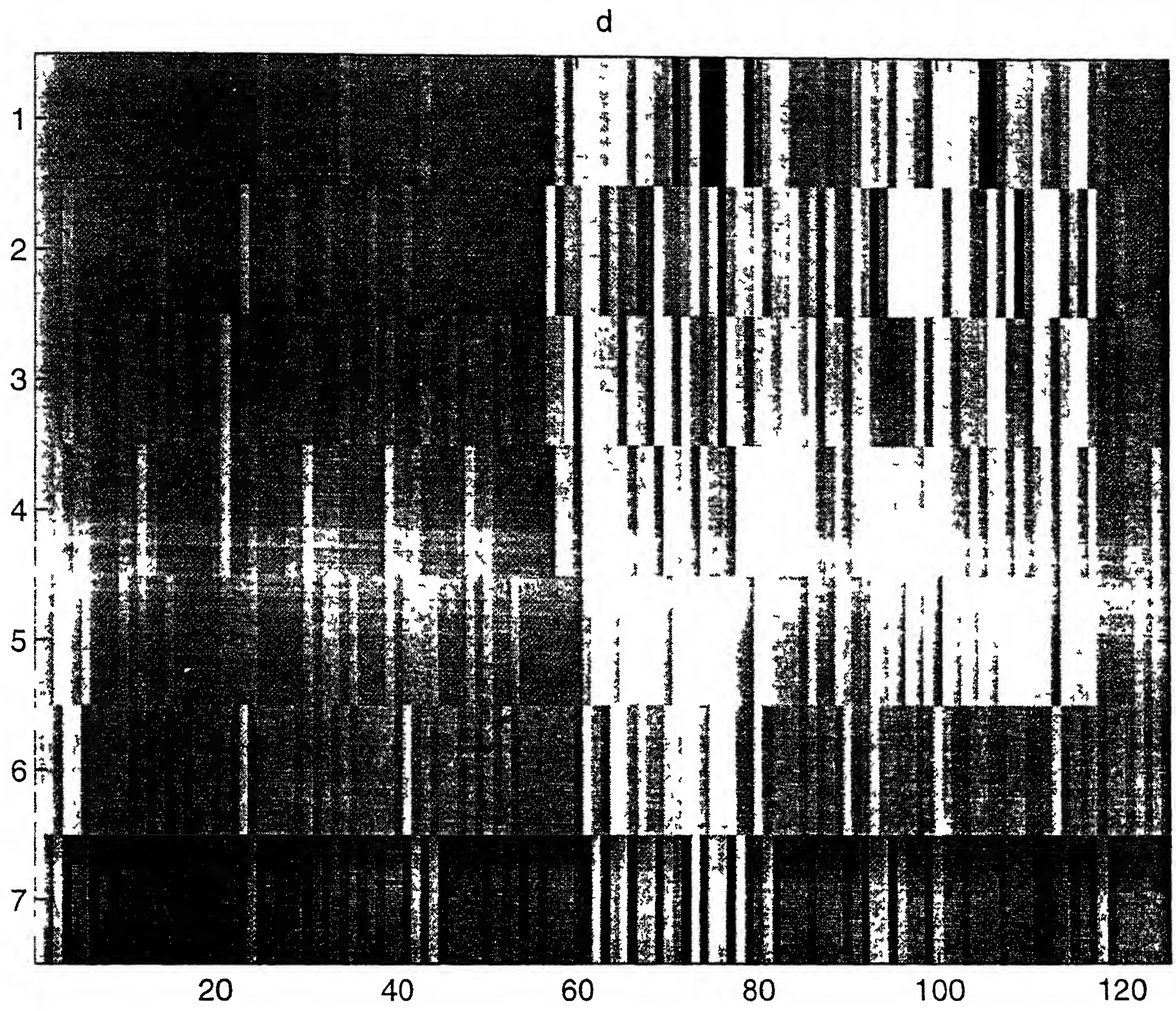

Figure 4 11 Modified Waveletogram of D

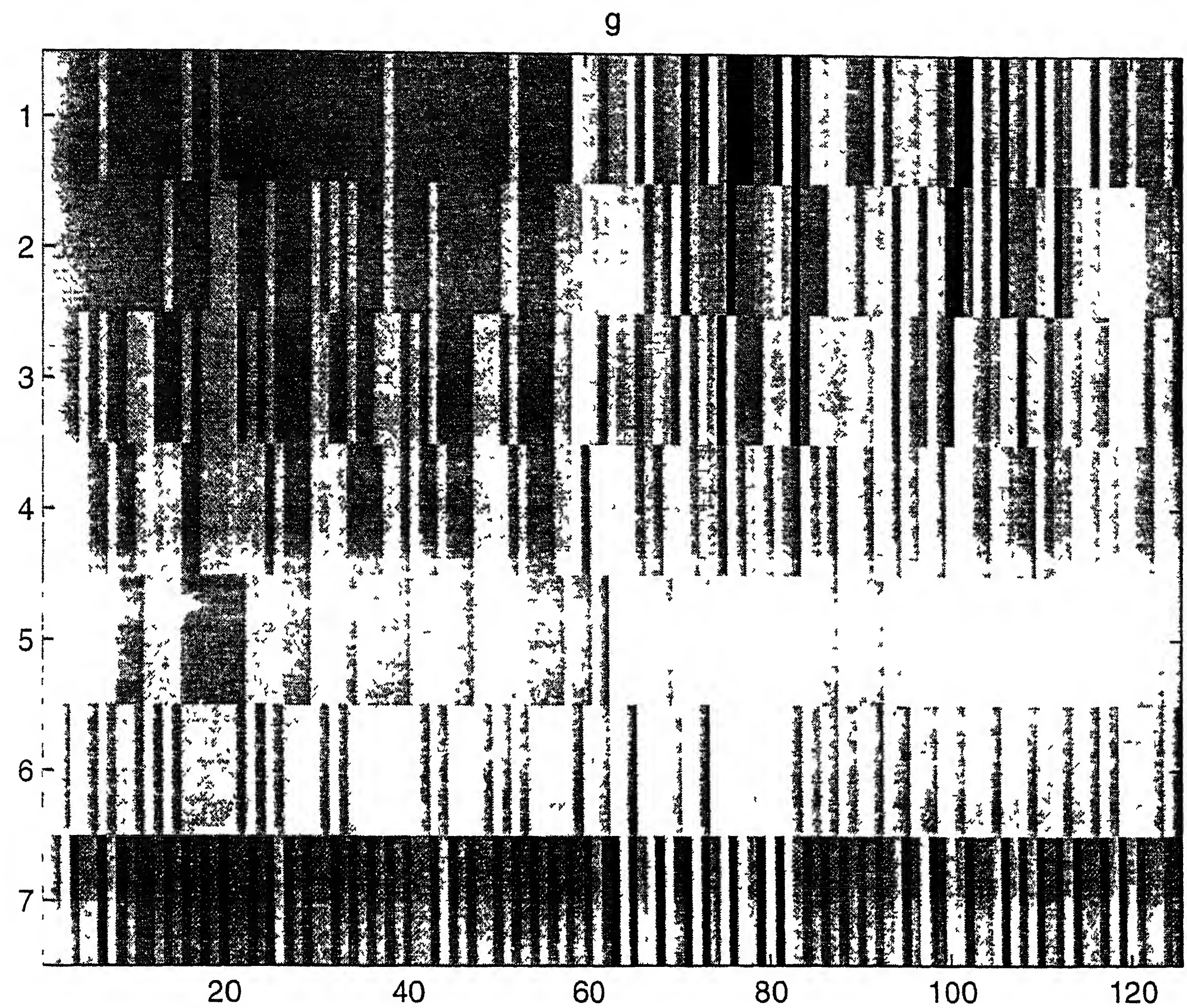

Figure 4 12 Modified Waveletogram of G

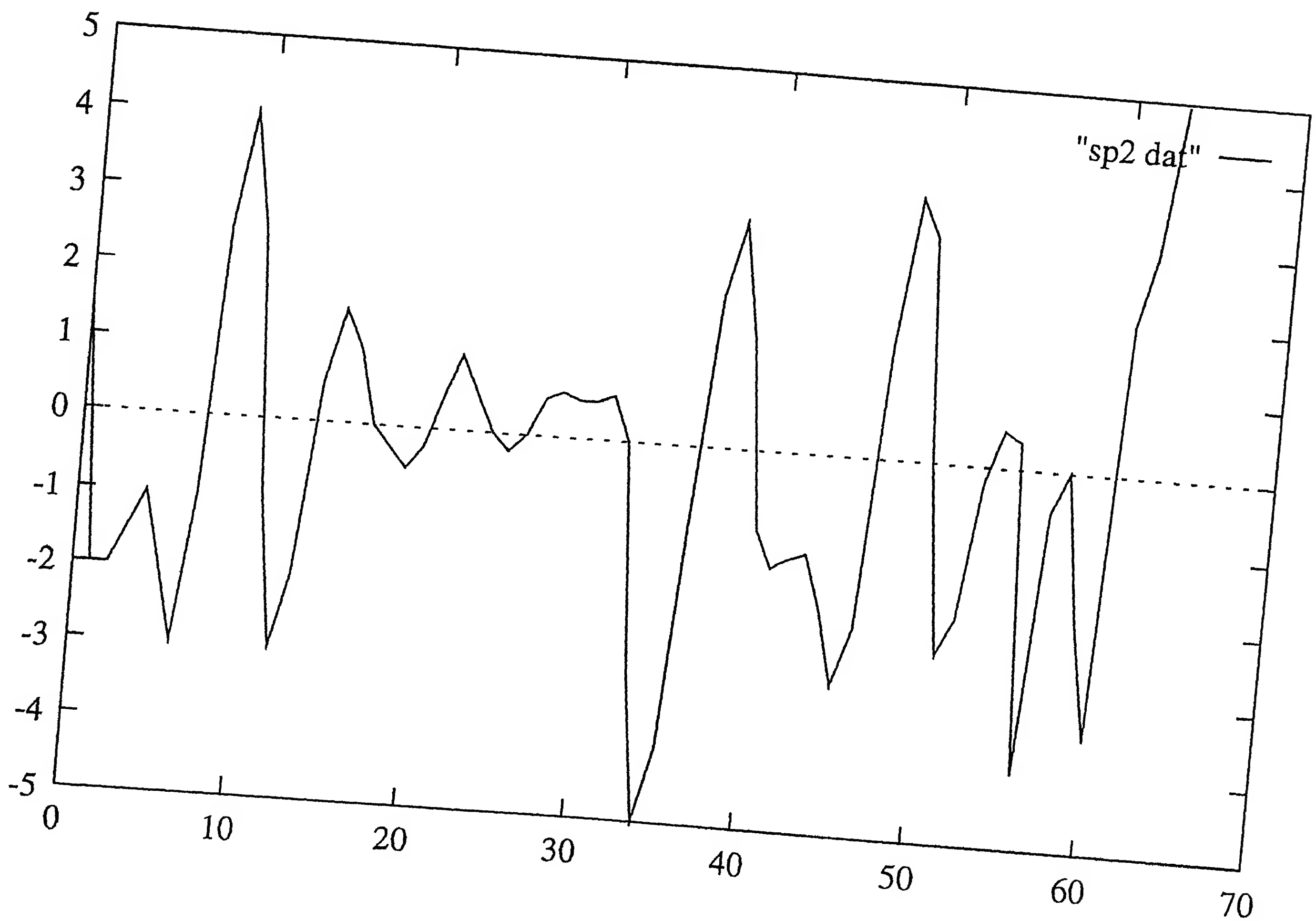

Figure 4.13 Original Signal

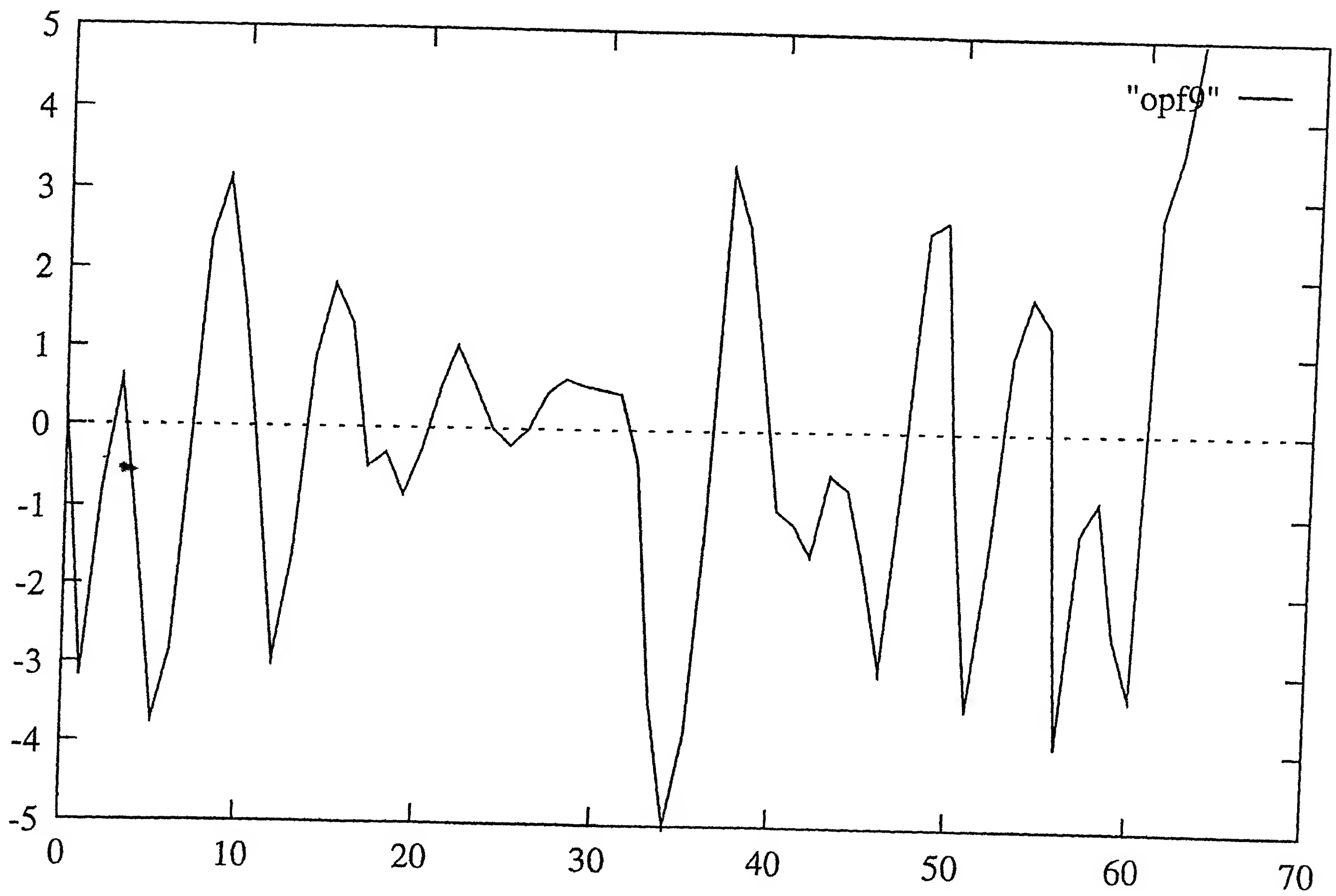

Figure 4 14 Signal Reconstructed after 20 Iterations

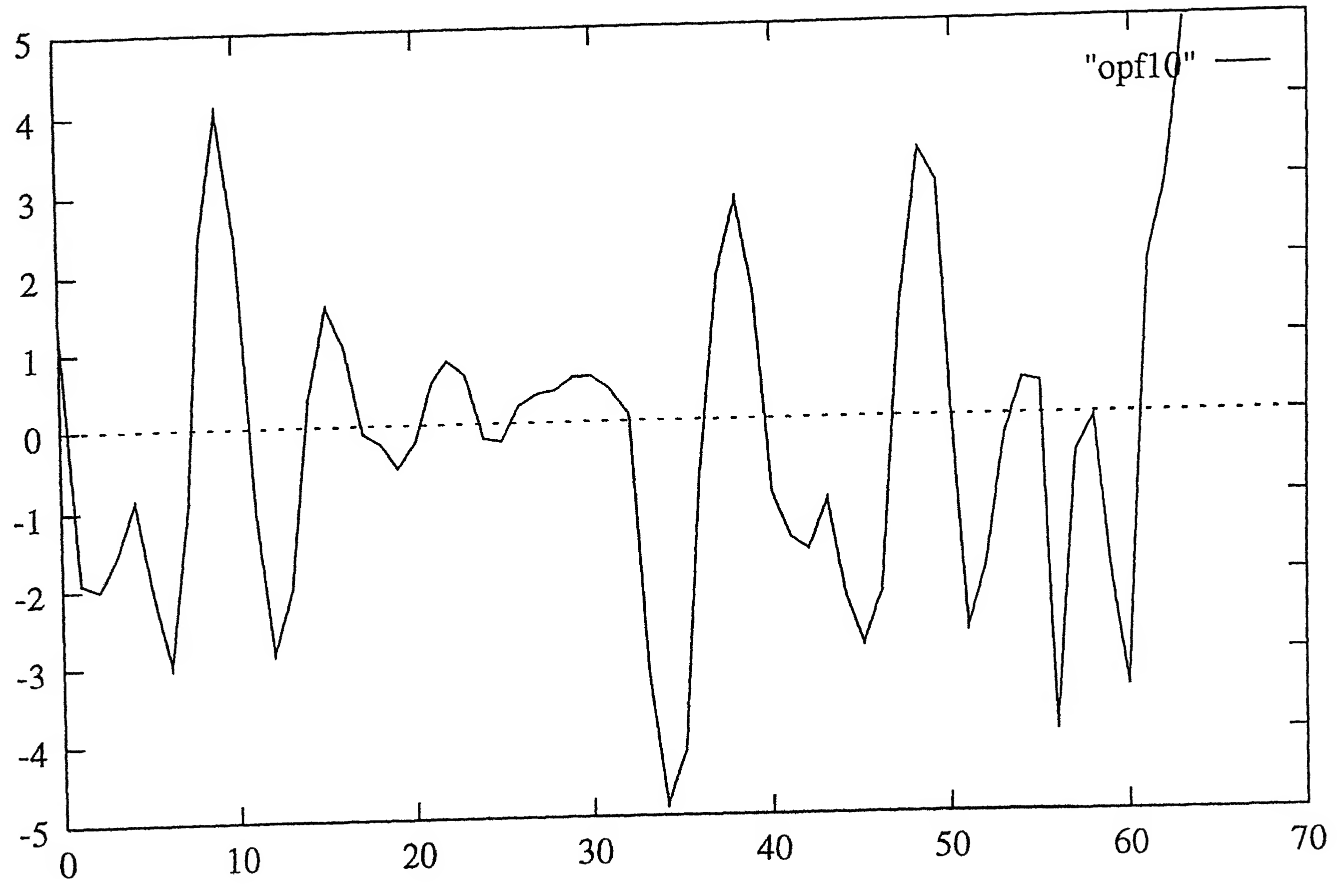

Figure 4 15 Signal Reconstructed after 50 Iterations

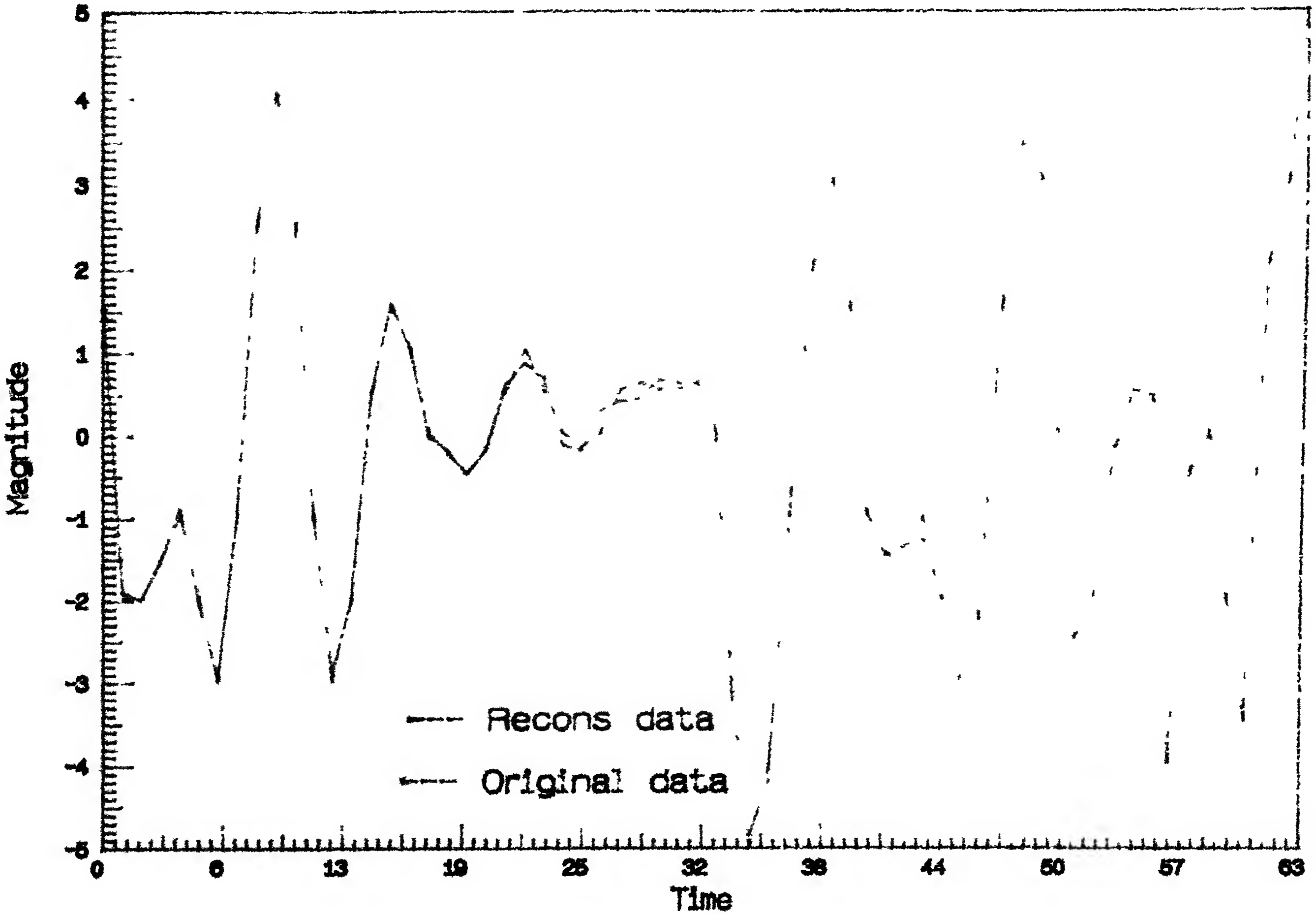

Figure 4.16: Comparison of Original Signal and Reconstructed signal after 50 Iterations

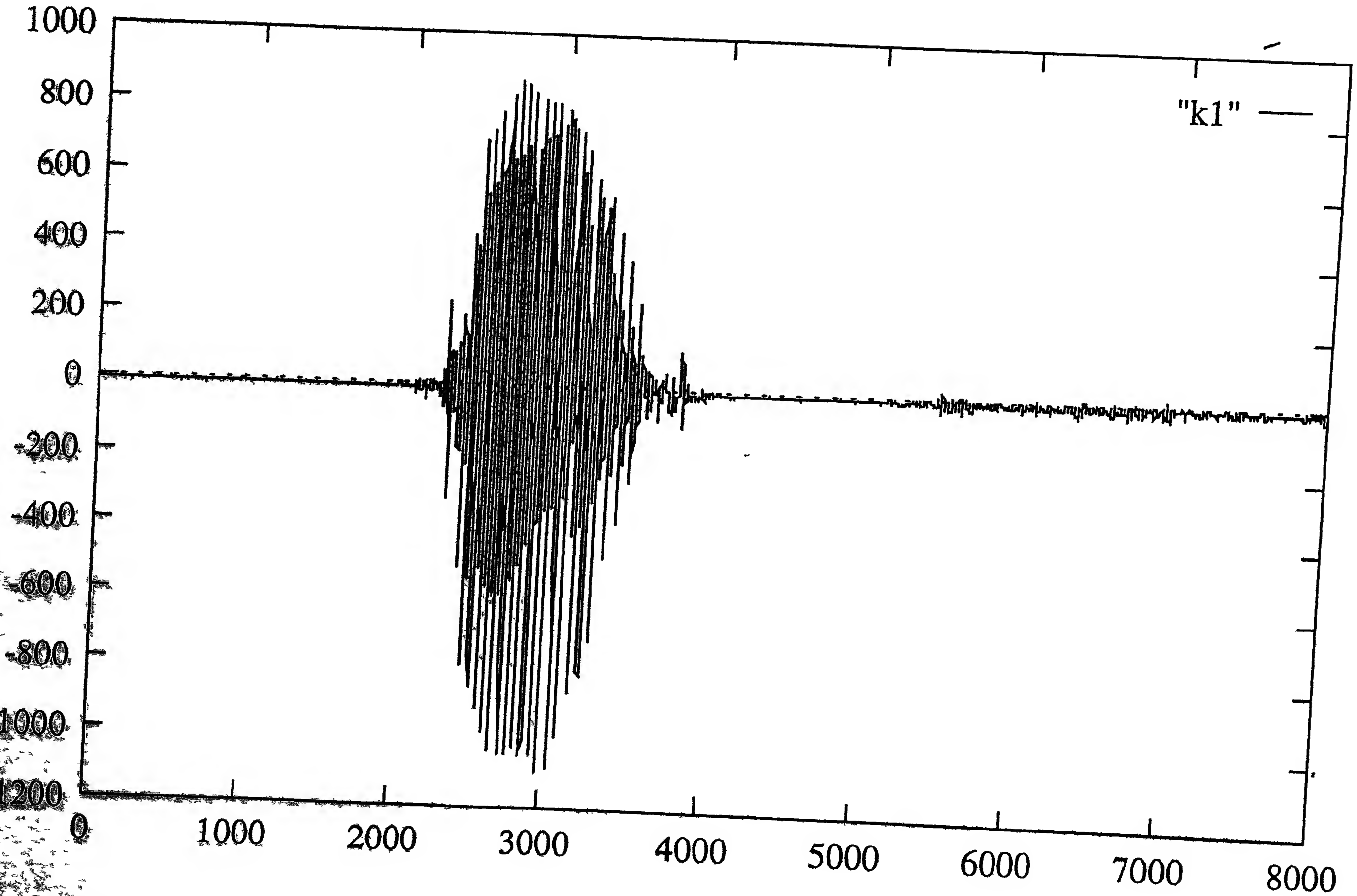

Figure 4 17 Time plot of /k/

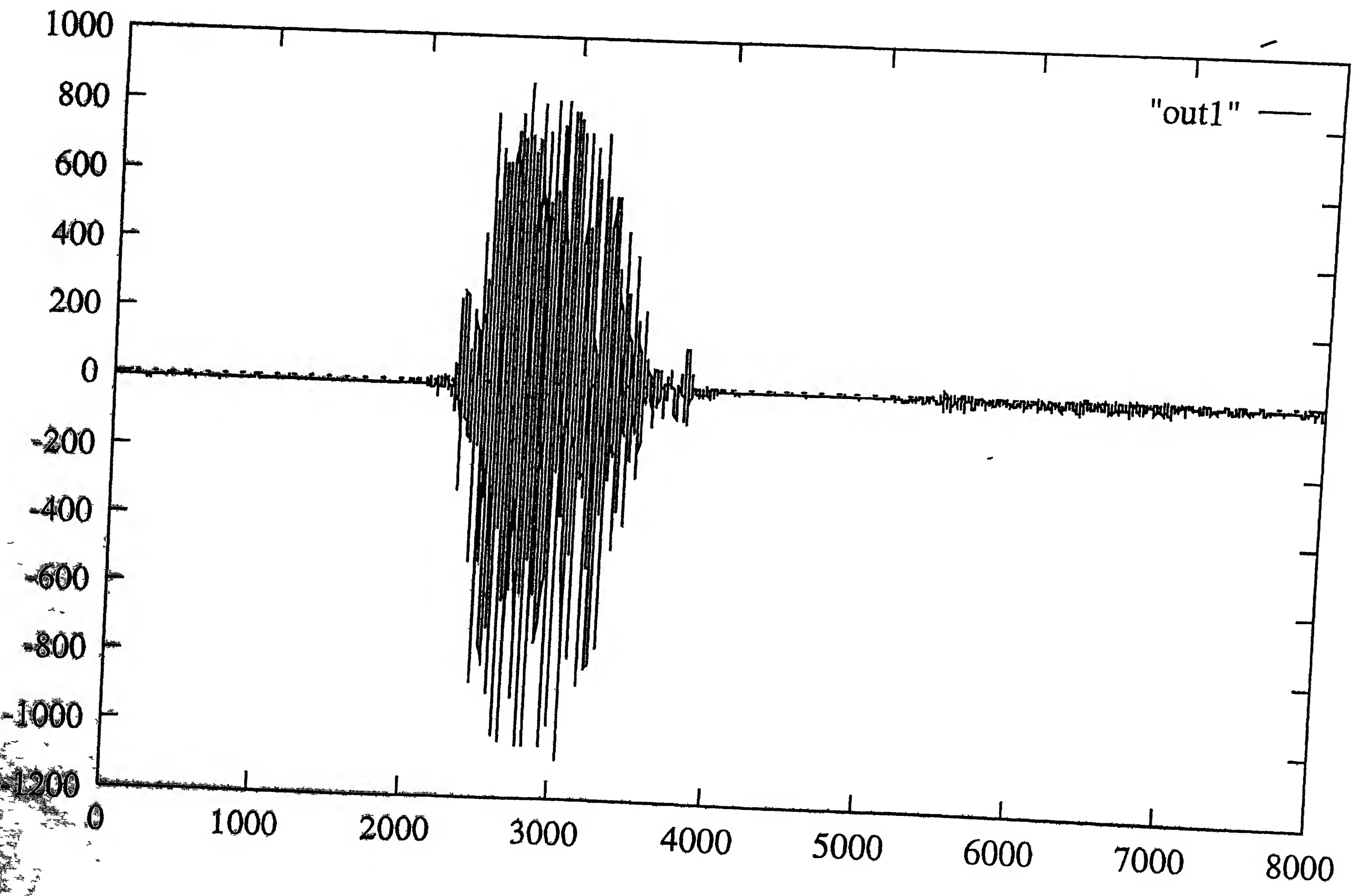

Figure 4.18: Signal Reconstructed after 20 iterations

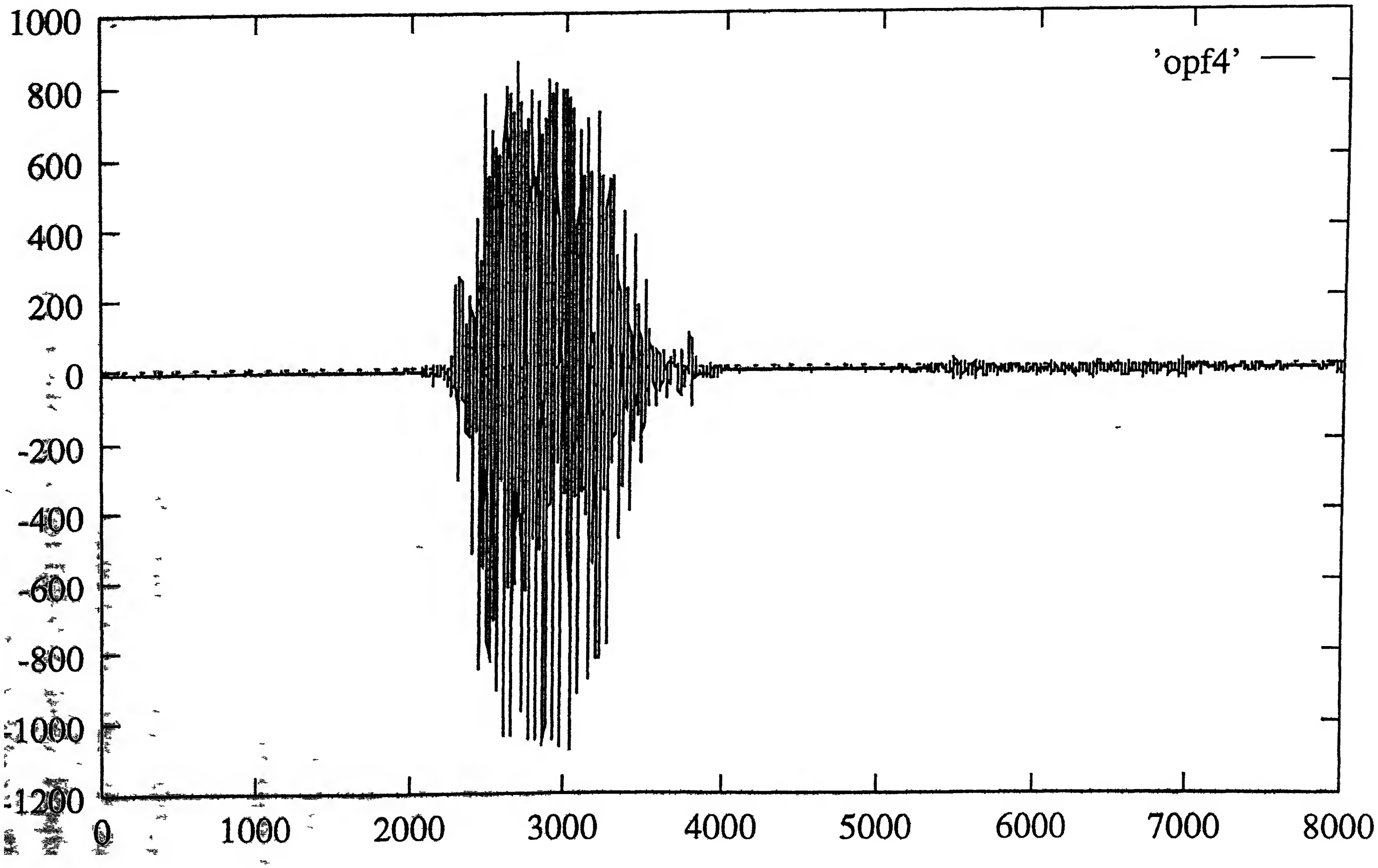

Figure 4.19: Signal Reconstructed after 50 iterations

Chapter 5

Conclusion and Suggestion for Future Work

Wavelet Transform has become a powerful tool in the area of signal processing The present thesis is an attempt to utilize this tool for analysis and characterization of stop consonants /k, p, t, b, d, g/ Three methods of analysis have been explored

Analysis through classification is based on k–means clustering method In this method 100 ms of speech signal was analyzed after the burst Wavelet transform was then applied on the resulting signal Five repetitions of each voiceless stop consonant /k, p, t/ were used in training data set and 10 parameters per levels were obtained through clustering method Five levels of wavelet transform were chosen Hence 50 parameters for detail signal and 10 parameters for coarser signal constitute 60 parameters per stop consonants Each stop consonant was then classified in terms of these stored parameters and 83% correct classification was achieved Learning Vector Quantization (LVQ) would have produced better results because the boundaries between different classes are optimized in this method instead of average distortion

Our effort to extract explicit time information like voice onset time (VOT), place of occurrence of burst etc was not successful In fact we tried to establish one–to–one correspondence between spectrogram and waveletogram, but failed to do so Frequency division on dyadic scale (in waveletogram) was the likely reason of this failure. Hence no explicit

time information could be obtained

Mallat's algorithm for reconstruction from modulus maxima of its wavelet transform is successfully implemented This algorithm shows that a signal can be well represented in terms of modulus maxima of its wavelet transform The reconstruction algorithm recovers a close approximation of sharp variation points The error is below our perception The algorithm was applied on stop consonant /k/ and a close replica was obtained as described in the previous chapter This suggests that stop consonant can possibly be described in terms of the modulus maxima of their wavelet transform at different levels or scales and their positions However, since the speech signal is highly oscillatory (see Fig 4 17), and in the wavelet transform domain, there are large number of local maxima, further data reduction (eg through thresholding) is needed before this characterization can be exploited This needs further investigation

In the present work, we have used Daubechies–10 wavelet on a dyadic grid for ease of computation As a hindsight, it is felt that another choice of wavelet (eg Butterworth[18]) and other possibilities (eg with scaling parameter $1 < a < 2$) would have been more appropriate

Adaptive frequency tiling [19] and computation complexity are other two important factors in the analysis Adaptive tiling is very important especially for waveletogram Computation complexity is a very important factor in real time application

Bibliography

[1] Y T Chan, "Wavelet Basics", *Kluwer Academic Press*, 1995

[2] Mallat S G, "A Theory for Multiresolution Signal Decomposition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol 11, No 7, pp 674-693,July 1989

[3] Mallat S G, "Multifrequency Channel Decomposition of Images and Wavelet Models", *IEEE Trans ASSP*, Vol 37, No 12, Dec 1989,pp 2091,2110

[4] Stephane Mallat and Sifen Zhong "Characterization of Signals from Multiscale Edges", *IEEE Trans PAMI*, Vol 14, No 7, july 1992,pp 710,732

[5] Daubechies I , "Ten Lectures on Wavelets", *Society for Industrial and Applied Mathematics*, 1992

[6] Daubechies I , "The Wavelet Transform , Time Frequency Localization and Signal Analysis", *IEEE Trans info Theory*, Vol 36,No 5 , pp 961-1005, Sept 1990 .

[7] Rioul O ,Vetterli M , "Wavelets and Signal Processing", *IEEE Signal Processing Magazine*, Vol 8,No 4 pp 14-38 , Dec 1990

[8] B Gidas and A Murua, "Classification and Clustering of Stop Consonants via Nonpara metric Transform and Wavelets", *IMA Proceedings* , 1995 ,Springer-verlag .

[9] B Gidas and A Murua, "Stop Consonants Discrimination and Clustering using Nonlinear Transformation and Wavelets", *Image Models* IMA Proceedings , 1994 ,Springer-verlag

[10] Rabiner L R , Schafer R W , "Digital Processing of Speech Signal", *Prentice Hall 1978*, Chapter 3

[11] D Youla and H Webb, "Image Restoration by the method of convex projections" *IEEE Trans Med Imaging* Vol MI-1,pp 81-101,Oct 1982

[12] Y Meyer , "Un contre-exemple a la conjecture de Marr et a celle de S Mallat," Reprint , 1991

[13] Y Linde etal , "Algorithm for Vector Quantizer Design" *IEEE Trans Comm* Vol COM-28,pp 84-95, Jan 1980

[14] I Daubechies,A Cohen and J C Feavueau ,' "Biorthogonal Basis of Compactly supported Wavelets" *Comm on Pure and Applied Mathematics* , Vol 45,No-5 , pp 485-560, June 1992

[15] R K Potter , G A Kopp , and H C Green , *Visible Speech* , Dover Publication Inc 1966

[16] T Cohonen , "Improved Version of LVQ" *IJCNN'90* , Vol 1, pp 545-550, June 1990

[17] R O Duda and P E Hart, "Pattern Classification and Scene Analysis" *New York* , Wiley , 1973

[18] Cormac Herley and Martin Vetterli , "Wavelets and Recursive Filter Banks" *IEEE Trans Signal Processing* , Vol 41,No-8 , pp 2516-2537, Aug 1993

[19] Cormac Herley Jalena Kovacevic, Kannan Ramachandran and Martin Vetterli , "Wavelets and Recursive Filter Banks" *IEEE Trans Signal Processing* , Vol 41,No-12 , Dec 1993

Appendix A

Projection Operator on Γ

In this appendix, we characterize the orthogonal projection on Γ in one dimensional space and also explained how to suppress oscillations The operator P_Γ transforms any sequence $(g_j(x))_{j\in Z} \in K$ into the closest sequence $(h_j(x))_{j\in Z} \in \Gamma$ with respect to the norm $\|\ \|$ Let $epsilon_j(x) = h_j(x) - g_j(x)$ Each function $h_j(x)$ is chosen so that

$$\sum_{j=-\infty}^{+\infty} ||\epsilon_j||^2 + 2^{2j}||\frac{d\epsilon_j}{dx}||^2 \tag{6 1}$$

is minimum To minimize this sum, we separately each component

$$||\epsilon_j||^2 + 2^{2j}||\frac{d\epsilon_j}{dx}||^2 \tag{6 2}$$

Let x_0 and x_1 be the abscissa of two consecutive modulus maxima of $Wf_{2^j}(x)$ Since $(h_j(x))_{j\in Z} \in \Gamma$, we have

$$\epsilon_j(x_0) = W_{2^j}f(x_0) - g_j(x_0) \tag{6.3}$$

$$\epsilon_j(x_1) = W_{2^j}f(x_1) - g_j(x_1) \tag{6 4}$$

Between the abscissa x_0 and x_1, the minimization of the (6 1) is equivalent to the minimization of

$$\int_{x_0}^{x_1}(|\epsilon_j|^2 + 2^{2j}|\frac{d\epsilon_j}{dx}|^2)dx \tag{6 5}$$

The Eular equation associated with this minimization is

$$\epsilon_j - 2^{2j}\frac{d^2\epsilon_j}{dx} = 0, \tag{6 6}$$

for $x \in [x_0, x_1]$ The constraints (6 3)& (6 4) are the boarder conditions of this membrane equation The solution is

$$\epsilon_j(x) = \alpha e^{(2^{-j}x)} + \beta e^{(-2^{-j}x)} \tag{6 7}$$

where the constraints α and β are adjusted to satisfy the constraints equations (6 3) & (6 4)

WE know that the modulus maxima of the original wavelet transform are only located at the positions x_n^j We can thus also impose sign constraints in order to suppress any spurious oscillation in the reconstructed wavelet transform This is done by imposing that the solution belongs to appropriate convex set y Let *sign(x)* be the sign of the real number x Let Y be the set of sequences $(g_j(x))_{j\in Z} \in K$ such that for any pair of consecutive maxima positions (x_n^j, x_{n+1}^j) and $x \in [x_n^j, x_{n+1}^j]$

$$sign(g_j(x)) = sign(x_n^j) \quad if \; sign(x_n^j) = sign(x_{n+1}^j) \tag{6.8}$$

$$sign(\frac{dg_j(x)}{dx}) = sign(x_{n+1}^j - x_n^j) \quad if \; sign(x_n^j) \neq sign(x_{n+1}^j) \tag{6.9}$$

The set Y is a closed convex and $(W_{2^j})_{j\in Z} \in Y$ Instead of minimizing $\|\;\|$ over $\Gamma \cap V$, we minimize it over $Y \cap \Gamma \cap V$ We thus alternate projections on Y, Γ, and V To compute the orthogonal projection on the convex Y,we need to solve an elastic membrane problem under constraints This can be done with an iterative algorithm that is computationally intensive Instead we implement a simpler projector P_Y on Y, which is not orthogonal with respect to the norm $\|\;\|$ Let $(g_j(x))_{j\in Z} \in K$ and $P_Y(g_j(x))_{j\in Z} = (h_j(x))_{j\in Z}$ For each index j, $h_j(x)$ is obtained by clipping of the oscillations of $g_j(x)$